Teneurin evolution and brain-specific functions

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Jan Beckmann

aus Marl, Deutschland

Basel, 2012

creative commons

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 2.5 Schweiz

---

Sie dürfen:

das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen

Zu den folgenden Bedingungen:

Namensnennung. Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).

Keine kommerzielle Nutzung. Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.

Keine Bearbeitung. Dieses Werk darf nicht bearbeitet oder in anderer Weise verändert werden.

- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen. Am Einfachsten ist es, einen Link auf diese Seite einzubinden.
- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.
- Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/ Datum: 3.4.2009

Summary

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Dr. Ruth Chiquet-Ehrismann

Richard P. Tucker, PhD

Dr. Jan Pielage

Basel, den 22.05.2012

Prof. Dr. Martin Spiess

Dekan

Summary

Table of Contents

1. Summary

Teneurins are a novel family of type II transmembrane proteins. They are highly conserved from invertebrates to vertebrates, in which four paralogs exist, called teneurin-1 to -4. Their main expression site is the developing nervous system, where they are expressed in distinct non-overlapping patterns. Further studies suggest that they have important functions in many different developmental processes, mainly at sites of migration and pattern formation. The exact mechanism of action of the teneurins is still under investigation. However, it does involve homophilic interactions and the release of the intracellular domain from the membrane and its translocation into the nucleus.

In the first part of the thesis, I am presenting a study of the possible evolutionary origins of teneurins, identifying a teneurin-like gene in the choanoflagellate *Monosiga brevicollis*. Furthermore, alternative splicing of the intracellular domain and conservation of nuclear localization sequences and protease cleavage sites were analyzed.

In the second part of the thesis, I analyzed the transcriptional regulation of human teneurin-1, contributing to its patterned expression in the brain. I found a novel conserved alternate promoter, upstream of the annotated transcription start. The transcription factor Emx2 directly binds to a single conserved homeobox binding site in this novel promoter, *in vitro* and *in vivo* and Emx2 is able to activate teneurin-1 expression in reporter assays, as well as on the endogenous level.

In the third part of the thesis, I investigated the homophilic interactions of teneurins, using an atomic force microscope (AFM) as a single cell force spectroscope (SCFS). This showed that the intracellular domain is essential for mediating adhesion forces, established by homophilic interactions by the extracellular domains. The NHL repeats of teneurins located in the large extracellular part between the EGF-like repeats and the YD-repeats are essential in discriminating homophilic versus heterophilic interactions. Finally, I could show that the homophilic interactions mediated by the NHL domains provide a signal for slowing down neurite outgrowth of Nb2a cells.

2. Introduction

Cell-cell interactions are fundamental for the regulation of many basic developmental processes and are mainly mediated by transmembrane proteins. The process of projecting neurons finding their target region in the brain is such an example, of how cell-cell adhesion contributes to regulating higher biological functions. Since the protein family analyzed, the teneurins, show characteristics of target recognition molecules, I will first introduce basic adhesion processes, followed by an introduction to the development of the nervous system and subsequently a more detailed introduction into the teneurin proteins.

2.1 Cell-Cell and Cell-ECM adhesion

In complex multicellular organisms, cells have to interact with each other and the surrounding extracellular matrix (ECM) to form and maintain tissues and organs. Figure 1 gives an overview of cell-cell and cell-ECM interactions. Mature cell junctions such as tight junctions, gap junctions, adherens junctions, desmosomes and hemidesmosomes have many different functions ranging from communication to providing stability. For these junctions to form and mature, cells have to adhere to each other first. Specialized integral membrane proteins called cell-adhesion molecules (CAMs) mediate these cell-cell adhesions. Many CAMs contain multiple distinct extracellular domains, which are usually present in repeats. Cell-cell adhesions are often mediated by two types of molecular interactions (Figure 2). In a first step, CAMs laterally associate on the same cell, called *cis* interactions, and form dimers or higher oligomers. By utilizing different domains, the CAMs can either bind *homophilic* to the same kind of CAM, or *heterophilic* to different CAMs expressed in an adjacent cell, called *trans* interactions. These two types of interactions can lead to a specific clustering of CAMs at contact sites, resulting in a high local concentration of CAMs. This is an important form of regulation of the adhesion force, because unlike other cell surface receptors, CAMs usually have a very low binding affinity to their ligands. The intracellular domains of CAMs can exert different types of functions. They can either act as signal transducer recruiting signaling complexes or as adapter proteins, directly or indirectly connecting CAMs to the cytoskeleton. The majority of CAMs can be categorized in four major protein families: cadherins, selectins, immunoglobulin (Ig) superfamily and integrins.[1]

The cadherins form with more than 100 members the largest group of CAMs. They can be further grouped into at least six subgroups. Cadherins have several different important

functions, particularly during early differentiation. The most widely expressed and best studied cadherins are the so-called “classical” cadherins: E- and P-cadherins expressed in epithelial cells and the neuronally expressed N-cadherin. All cadherin interactions are dependent on the extracellular concentration of Ca^{2+}-ions. All cadherins have highly conserved intracellular domains, mediating interactions with the cytoskeleton in order to establish the adhesive forces of cadherin-dependent cell adhesion.[2] Many cadherins and protocadherins are expressed in distinct patterns in the developing and adult nervous system and show concentrated localization at synapses. It is believed that cadherins do not only establish embryonic structures, but also functional compartmentalization and synaptic connectivity.[3]

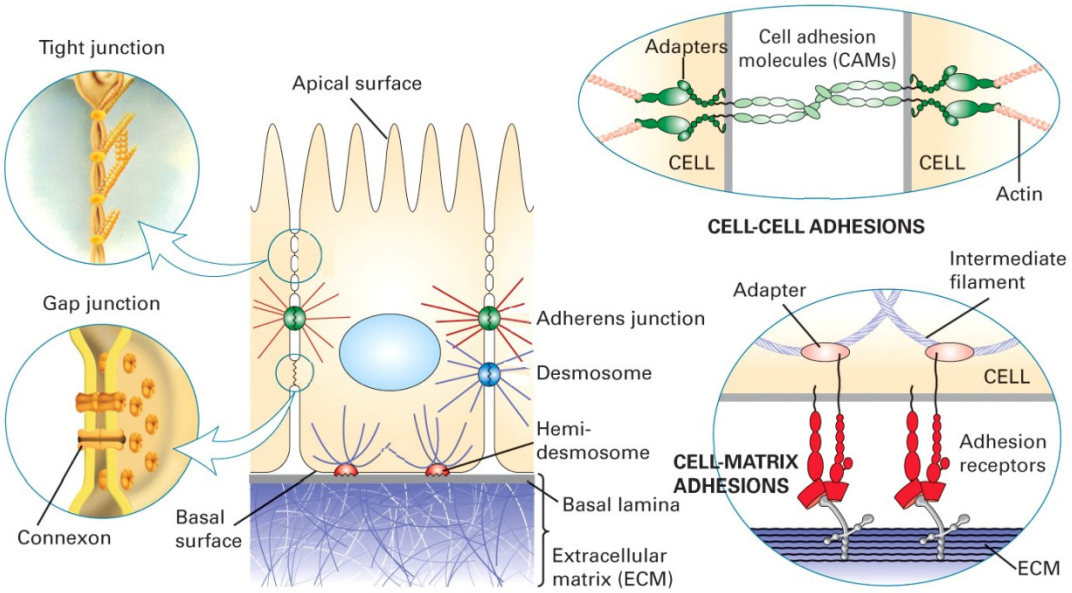


Figure 1: Schematic overview of major adhesive cell-cell and cell-ECM interactions [1]
Tight junctions, Gap junctions, adherens junctions, Desmosomes and Hemi-Desmosomes are various mature cell junctions that act as important mediators of tissue integrity and communications between cells. Cell-cell adhesion is mediated by cell adhesion molecules (CAMs).

Another Ca^{2+}-dependent class of adhesion molecules is formed by the selectins. Selectins bind to carbohydrates at the cell surface. The dominant sites of expression are blood and endothelial cells. They play an important role in leukocyte extravasation from the blood stream, as well as in other inflammatory processes.[4]

Introduction

The Immunoglobulin-like CAM (Ig-CAM) superfamily mediates cell adhesion independent of Ca^{2+}-ions. The typical Ig domains consist of two anti-parallel β-sheets, which form a sandwich.[5] The number of these modules varies. The first and best described member of this family is NCAM, which is widely expressed in most cell types, including most neurons. NCAMs exist in three major isoforms, but at least 20 isoforms are generated by alternative splicing. All NCAM isoforms can be heavily modified posttranslationally. Just like other CAMs, NCAMs can increase cell adhesion by homophilic binding, but they were also shown to bind heterophilically to other partners and perform signaling functions.[6]

Integrins perform different functions from the other CAM families described above. While some integrins are involved in cell-cell interactions, the majority of integrins are ECM receptors. Integrins are glycoproteins that form non-covalently linked heterodimers, consisting of an α- and β-subunit. In mammals, 18 α- and 8 β-subunits are encoded in the genome, which can combine to a total of 24 αβ-integrin receptors. Most of the ECM proteins like collagens, fibronectin and laminin are recognized by specific integrins. Many integrins can be found in different conformational states, influencing their receptor affinity. An important regulator of integrin-mediated adhesion is the clustering of the receptors at contact sites.[7] Like the other CAM families, integrins do not only function in promoting cell adhesion, but also activate intracellular pathways leading to proliferation, survival and migration.

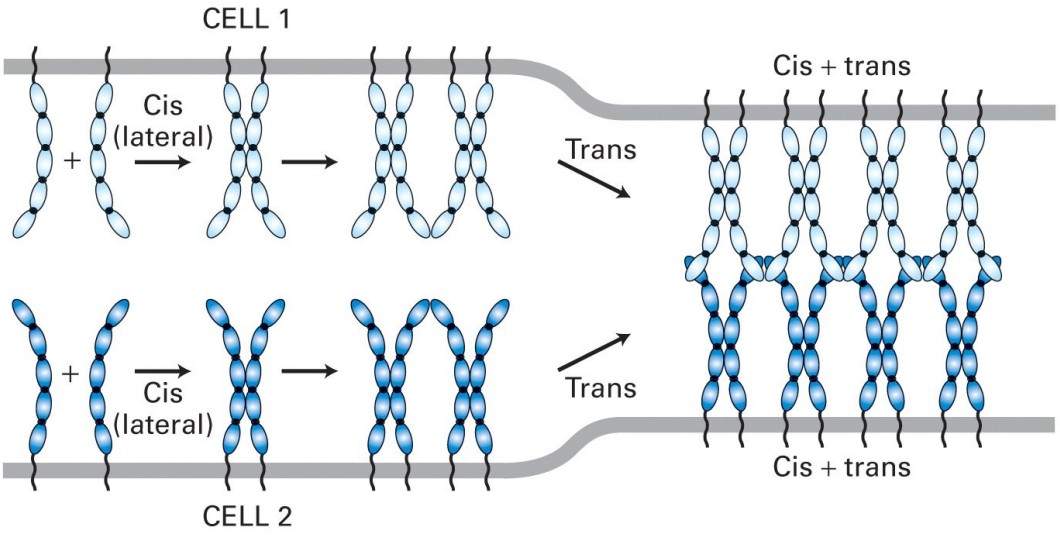


Figure 2: Schematic model for the generation of cell-cell adhesions [1]
Within the plasma membrane of one cell, CAMs form dimers and higher oligomers. This interaction on the same cell is called "*cis*"-interaction. Interactions of different domains of the CAMs on adjacent cells generate strong adhesion forces between two cells and are called "*trans*"-interactions.

2.2. Early neuronal development and axon guidance

2.2.1 Early neuronal development

During gastrulation, the primitive two-layered embryo is transformed into a three-layered structure. The epiblast cells of the upper layer will differentiate into the three primary germ cell layers: endoderm, mesoderm and ectoderm. The ectoderm in vertebrates is divided into three parts: the external ectoderm, which will give rise to the epidermis of the skin, the neural crest and the neural tube, which together form the neuroectoderm. Therefore, the entire nervous system of vertebrates is derived from the ectoderm.

After the neural progenitor cells are produced during gastrulation and positioned along the rostral-caudal midline, the formation of the neural tube as the first defined neural structure is the next major step in brain development.

During the first step of neural tube formation in vertebrates, two ridges form on both sides of the midline, where the neural progenitor cells are located. These ridges grow, fold inward and form a hollow tube. The closure of this tube begins in the center of the embryo, while the caudal end is the last to close. The neural tube is thus filled with a single layer of neural progenitors adjacent to the hollow center. This zone is called the ventricular zone. The neural tube undergoes rapid growth and eventually forms the three primary brain vesicles (Figure 3A). The most anterior vesicle, the prosencephalon is the precursor of the forebrain. The mesencephalon in the center will give rise to midbrain structures, whereas the most posterior rhombencephalon will become the hindbrain. These primary vesicles further subdivide and form the five secondary brain vesicles (Figure 3B). The primary organization of the central nervous system is now established.[8, 9]

Introduction

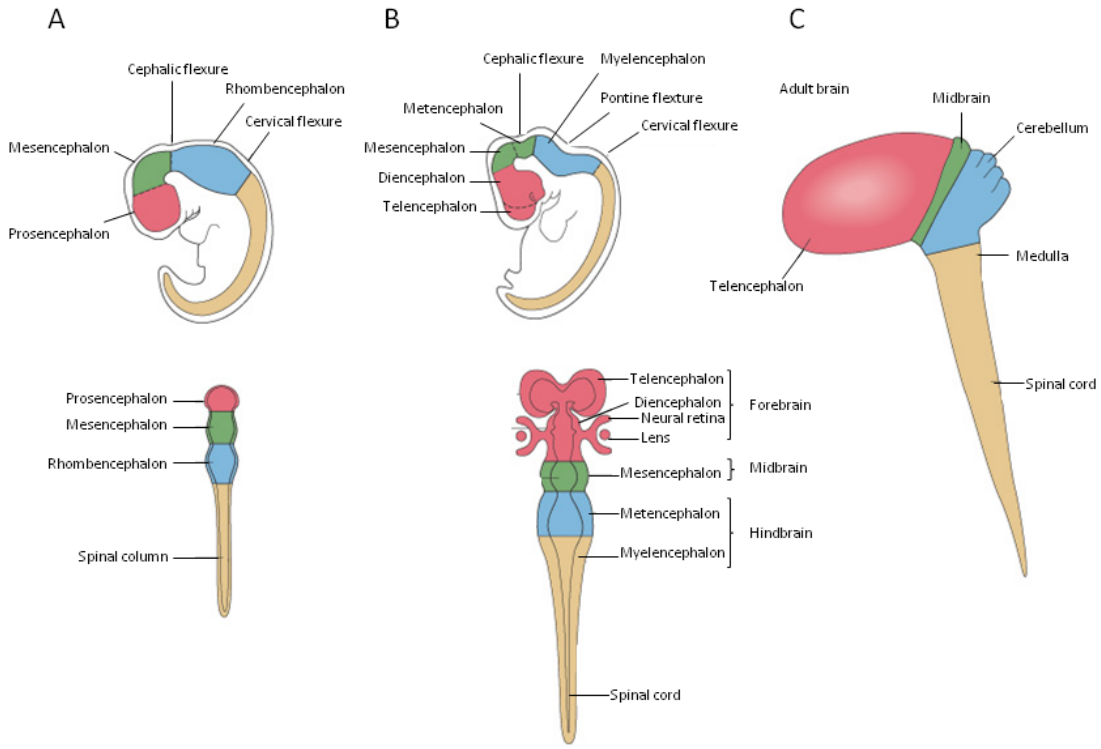


Figure 3: Schematic representation of vertebrate brain development and segmentation
Lateral (upper panel) and dorsal views (lower panel) of successive older stages of vertebrate development are shown (A,B,C) (A) The primary three brain vesicles are shown. (B) The primary three vesicles are further subdivided into five secondary vesicles. (C) The basic brain segmentation can be related to the overall organization of the mature brain (adapted from [10])

2.2.2 Patterning and segmentation of the brain

Segmentation of the brain into different vesicles and those vesicles into neuromers is a key concept in brain development. The further division into functional subunits (nuclei) is called patterning. These nuclei have distinct functions and patterns of connectivity (see Figure 4).[11] During the development of the brain, this specialized arealization is set up by a complex interplay of networks of transcription factors and secreted signals.[12] This organization is best studied in the cerebral cortex.

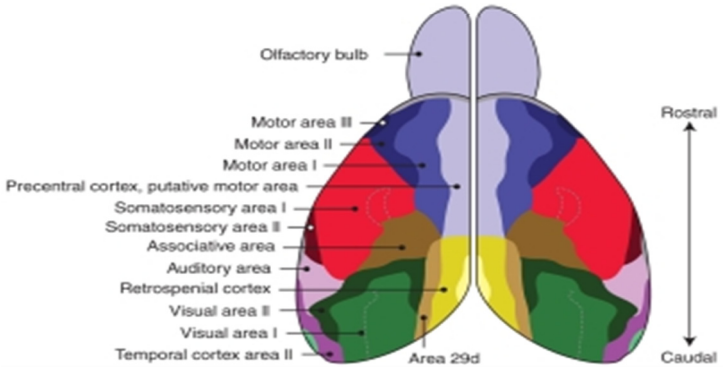


Figure 4: Dorsal view of an adult mouse brain[13]
The adult mouse brain is organized into functionally and histologically distinct areas. Different major areas are labeled. The motor area is located in the rostral part, whereas the visual areas are located caudal. The somatosensory areas are located in between, in the medial part.

The described primary organization of the brain is a result of neural patterning in the neural plate. This anterior-posterior patterning of the brain resulting in the five secondary vesicles is directed by a variety of proteins.[14] Most prominent among those factors are the *hox* genes. The homeobox domain-containing transcription factors are expressed in an ordered pattern in different tissues, with the brain being the major expression site. *Hox* genes are homologs of the Drosophila *Hom-C* genes, which are important for segmentation of the body axis. In the vertebrate brain, the defined expression of a combination of different *hox* genes defines the position and identity of the rhombomeres in the rhombencephalon, the developing hindbrain.[15]

The dorsal-ventral polarity is mostly established by secreted signals. These signals are produced in signaling centers along the edges and midline of the neural plate. The sonic hedgehog (Shh) protein is expressed in the ventral region of the neural plate.[16] On the dorsal side, secreted bone morphogenetic proteins (BMPs) regulate dorsal cell identities.[17] Another

secreted protein important for arealization in the developing brain is the fibroblast growth factor 8 (Fgf8).[18] Together with the signal molecules mentioned above, it forms a complex network of regulation. Changes in the expression strength of each member of this network by either a gene knock-out or overexpression, will lead to alterations in brain morphogenesis.[19]

The secreted factors mainly establish the specialized area identity through the graded expression of transcription factors. Several transcription factors are expressed in gradients along the dorso-ventral and rostro-caudal axes of the cerebral cortex. The most prominent members are COUPTF1, Lef1, Lhx2, Emx2 and Pax6.[11] The latter two factors are well-studied and play an important role in the early patterning of the cortex.[20] The homeodomain transcription factor Emx2 is expressed in a low rostral to a high caudal gradient.[21] The paired-box containing transcription factor Pax6 is expressed in an opposite gradient to Emx2 (see Figure 5 A).[22] The balance of expression of those two transcription factors regulates the functional fate of a given area along the concentration gradient. High concentrations of Emx2 combined with a low concentration of Pax6 lead to the production of neurons of the visual cortex, while the opposite concentration gradient, leads to motor cortex formation. Studies of mutant mice for both of these transcription factors show systematic changes in size and positioning of functional cortical areas.[21, 23, 24] In the *Emx2* knock-out mouse, the somatosensory and motor areas grow in size, while the visual areas shrink. In *Pax6* mutant mice the opposite effect is observed. The visual areas become enlarged, while the motor and somatosensory areas shrink (see Figure 5 B). Emx2 and Pax6 do not act alone in setting up functional patterning of the brain. Increasing evidence shows that they are part of a complex network of transcription factors, in which the relative concentrations to each other define the final functionality of the individual brain areas.[13] Downstream of these transcription factors are genes coding for different families of proteins, including other transcription factors, adhesion molecules and secreted axon guidance molecules. These protein families contribute to specifying the function and axonal connectivity of the brain area, in which they are expressed.[11]

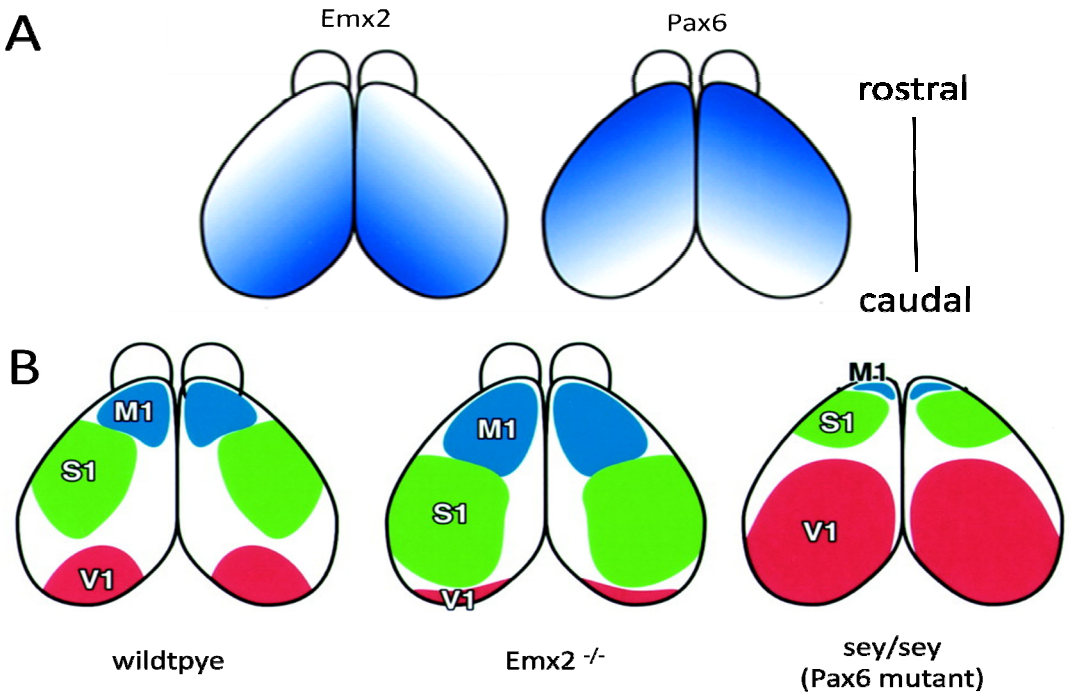


Figure 5: Diagrams of dorsal view of the mouse neocortex

(A) Graded expression patterns of Emx2 and Pax6 across the embryonic cortex. Emx2 is expressed in a low rostral to a high caudal gradient. Pax6 is expressed in an opposite gradient to Emx2. (B) Schematic representation of the organization of the cortex in different functional areas. The organization of those areas changes disproportionately in size and location in the different mutants. (V1) primary visual area; (S1) somatosensory area; (M1) motor area (adapted from [23])

2.2.3 Axon guidance and synaptogenesis

The nervous system depends on correct patterns of neuronal connectivity to function properly. How arealization and functional patterns are regulated at the transcription level is described above. To connect these areas, projecting neurons have to find their way and reliably make connections to their appropriate synaptic partners in their target areas. Responsible for the navigation of an axon is a specialized structure on the distal tip, called the growth cone. The growth cone is a highly motile structure, which constantly integrates inputs from the environment and thus controls the rate and direction of axonal growth. The guidance cues integrated by the growth cone can be either repulsive or attractive, as well as short or long ranged (see Figure 6).[25] The growth cone has to react to the appropriate set of cues and select the correct path to its target, which can be located at a large distance away. It does so, with specialized sets of receptors on its cell surface activating downstream signaling pathways and changing the cytoskeletal organization.[26] Members of four different ligand families are frequently found to be guidance cues in genetic and biochemical assays. There are slits, netrins, semaphorins and ephrins.

Slits are large secreted proteins, with their correspondent receptors on the growth cone, the "roundabout" receptors (Robos).[27] In *Drosophila*, slit is expressed in the ventral midline, where it prevents ipsilateral axons from crossing and commissural axons from recrossing.[28] In vertebrates, slits are essential for the formation of the optic chiasm.[29]

Netrins are secreted proteins that attract axons toward the midline, but also repel another subset of axons. Their corresponding receptors on the growth cone surface are the "deleted in colorectal carcinoma" (DCC) receptors.[30] Netrins can either function over long ranges up to a few millimeters or act as a short range guidance cues on other neurons. The diffusion range of netrins is still under debate.[31, 32]

Semaphorins are a large family of membrane bound and secreted guidance cues. There are eight structural classes, all sharing a conserved Sema domain at the N-terminal.[33] The corresponding receptors are multimeric. A protein of the plexin family is often part of the receptor complex. Several studies show that semaphorins primarily act as short range repulsive cues that deflect axons away from inappropriate targets. However, some studies suggest that semaphorins can also act as attractive molecules for a subset of neurons.[34]

Introduction

There are two classes of ephrins (Eph) and their receptor tyrosine kinase Eph receptor with over a dozen members. Ephrin-As are glycosylphosphatidylinositol anchored and bind to Eph-A receptors, whereas ephrin-Bs have a transmembrane domain and bind to Eph-B receptors.[35] Many of the ephrins function as a contact repellent thereby mediating axon fasciculation. In the visual system, ephrins and their receptors are expressed in a complementary gradient. The topographic position of retinal axons along the anterior-posterior axis is regulated by the ratio of the ephrin ligand to its receptor.[36]

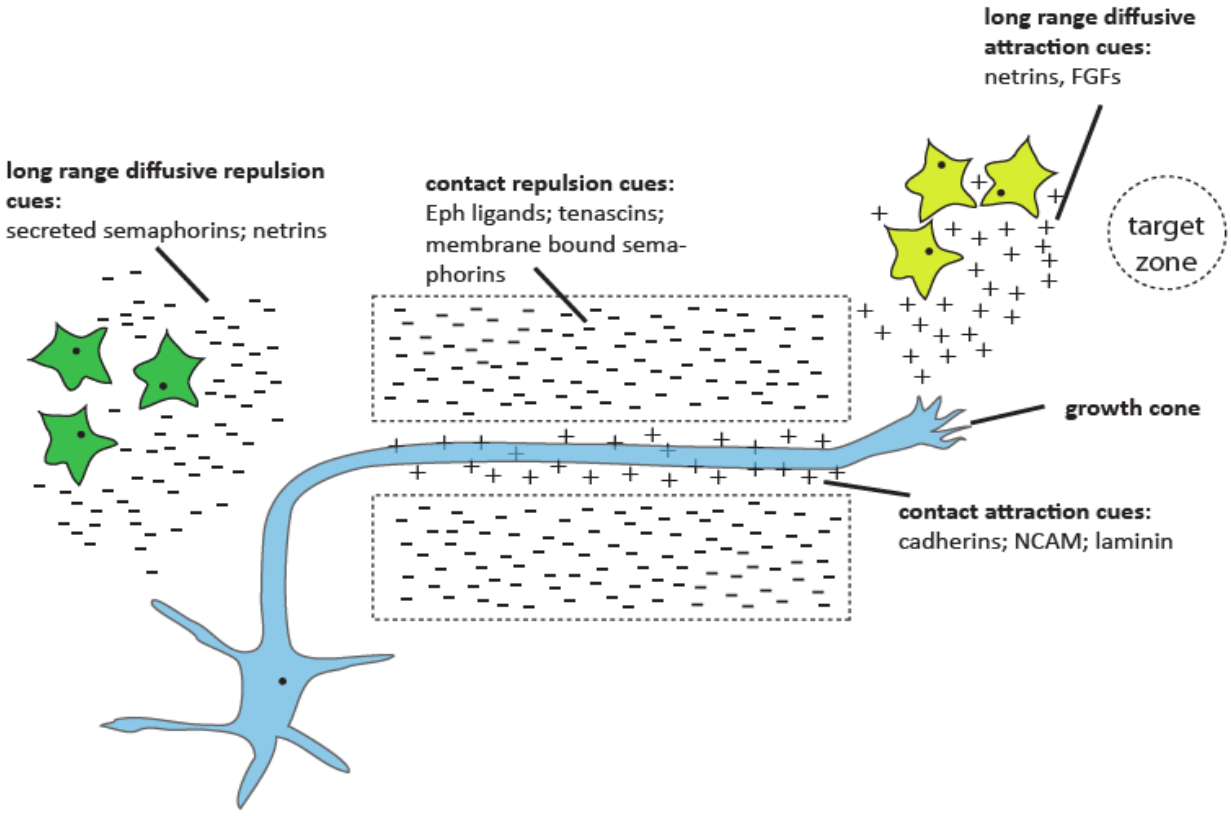


Figure 6: Schematic representation of axon guidance
Projecting neurons are guided to their target areas by guidance molecules. Those molecules can either be secreted and act long range or they are membrane bound and act short range as contact cues. Furthermore, they can act either repulsive or attractive for a single axon. The information of those molecules is integrated by receptors located in the growing tip of the axon called the growth cone. (adapted from [25])

The described main functions of the different classes of guidance cues are just a fraction of the known functions of these molecules. Furthermore, some molecules can have opposite effects on different subpopulation of approaching neurons. Additionally, the same growth cone can react to the same guidance molecule in a different fashion, depending on the location along the path toward its target. This complicates the analysis of specific functions for these molecules. However, this heterogeneity is an important characteristic of axonal guidance molecules, in order to ensure the correct wiring of the incredibly large number of neuronal connections, while minimizing the number of different molecules needed.

Additionally to these guidance molecules, CAMs like NCAM and N-cadherin are also involved in axonal guidance.[5, 37] CAMs usually direct axon guidance by promoting neurite outgrowth. As described in chapter 2.1. most CAMs also have a signaling function, which contributes to direct axon guidance. Furthermore, ECM proteins have been shown to contribute to axonal pathfinding. Laminin and fibronectin, for example, promote neurite outgrowth and the growth cone accurately follows pathways were laminin or fibronectin are deposited.[38] In contrast, growing neurons are deflected from areas were tenascins are deposited.[39]

Once a projecting neuron reaches its target area, it forms synapses with its appropriate partner neuron. This target finding is an important aspect of synaptogenesis, the process to form mature synapses. Invading neurons have to be activated or “primed” to form synapses. Target or glial cell derived soluble factors, like Wnts or FGFs promote neuronal maturation to make neurons competent to form synapses.[40, 41] As described for axonal guidance, several classes of CAMs have been shown to carry out functions in target recognition and initial formation of synapses. The cadherins are localized at synapses at early stages of synapse formation.[42] However, different studies have shown that cadherins are important for correct target specification, rather than the induction of synapse formation.[43, 44] Protocadherins are related to cadherins and studies suggest that they are also involved in target finding rather than synapse formation.[45] Different molecules are needed for inducing various aspects of synapse formation. Among them are neuronal activity-regulated pentraxin (Narp), Ephrin-B1 and the two CAMs, SynCAM and neuroligin. Narp was one of the first molecules, found to have synaptogenic activity and it was shown to cluster glutamate receptors in spinal cord neurons in mixed co-culture assays.[46] Ephrin-B1 belongs to the ephrin family of axonal guidance molecules. In synaptogenesis, ephrin-B1 clusters N-methyl-D-aspartate (NMDA)

types of glutamate receptors and thus contributes to post-synaptic organization.[47] SynCAM and neuroligin have a more general function in differentiating the presynaptic active zone. SynCAM is a member of the Ig superfamiliy of CAMs and interacts in a homophilic manner. Cultured neurons overexpressing SynCAM1 show an increase in synapse formation. Expression of SynCAM1 in non-neuronal cells induces functional presynaptic active zone formation in contacting axons in a mixed-culture assay.[48] Like SynCAM, neuroligin and its presynaptic partner neurexin induce the formation of the functional active zone by clustering.[41] These steps describe the early stages of synaptogenesis, in which CAMs and axonal guidance molecules play significant roles. Subsequently, the synapses mature by recruitment of other pre- and postsynaptic molecules, they expand in size and the numbers of synaptic vesicles increase substantially (for more details see [49]).

2.3. Teneurins

2.3.1. Conserved teneurin domain structure

Teneurins are a family of proteins, conserved from *C.elegans* and *Drosophila* to vertebrates.[50] They were first discovered in *Drosophila*, using different approaches in two independent laboratories. In the first approach, a screen was performed to identify homologs of the extracellular matrix protein tenascin-C in arthropods. Screening a *Drosophila* expression library with a probe against the epidermal growth factor-like (EGF-like) repeats of chicken tenascin-C identified two novel proteins that they named "tenascin-like protein major" (ten-m) and "tenascin-like protein accessory" (ten-a).[51, 52] More detailed studies showed that they had discovered a new class of proteins. Beside the EGF-like domain, the newly discovered proteins were distinct from tenascins in both, structure and function. Independently, ten-m was also found in a screen for identifying phosphotyrosine containing proteins.[53] Embryos, containing a mutant version of this gene showed an "oddless" pair-rule phenotype and hence the newly identified protein was named "odd Oz" (odz). Since all other known pair-rule genes code for transcription factors, this phenotype was very surprising for a transmembrane protein. Only recently it was shown, that the pair-rule phenotype was rather due to a problem of the balancer chromosome used in these mutants and not caused by the identified gene.[54] Further studies revealed the conservation of these proteins in different classes of organisms. In vertebrates, four paralogs were identified, whereas only one gene was found in *C.elegans* that is transcribed and translated in two isoforms.[55] As both names persisted being used, the search for relevant literature was complicated. Therefore, the new name "teneurin" was proposed.[56] It consists of the original name "tenascin- like protein" and its affiliation with the nervous system, the major site of expression. Fortunately, the numbering schemes used together with the different nomenclatures coincided. The name teneurin is now commonly used (see Table 1 for nomenclature overview).

All teneurins share a common domain structure. They are type II transmembrane proteins with a molecular mass of around 300 kDa. Teneurins consist of a short intracellular domain (ICD) and a very well conserved large extracellular domain (ECD). Within vertebrates, the ICD is also highly conserved. It contains two EF-hand-like calcium-binding domains, proline-rich stretches that are characteristic for SH3-binding sites and putative phosphorylation sites.

Table 1: Summary of nomenclature and expression sites during development (includes only proteins with known expression patterns)

Species	Name	Synonyms	Expression pattern
Caenorhabditis elegans	Ten-1L[55]		Somatic gonad, vulva, subset of neurons, gut, some hypodermal and muscle cells[55], pharyngeal and intestine cells[57]
	Ten-1S[55]		Subset of neurons, some hypodermal cells[55],[57]
Drosophila melanogaster	Ten-m[51, 56]	Odz[53, 58]	Odd-numbered parasegments[52, 53], subset of neurons, tracheal system, cardiac cells, lymph glands, muscle attachment sites[52], morphogenetic furrow[56, 58], wing pouch, leg and antennal discs[58], motor neuron[54]
	Ten-a[52, 59]		CNS, eye, muscle attachment sites[51, 59]
Danio rerio	Teneurin-3	Ten-m3[60]	Developing brain, somites, notochord, pharyngeal arches[60]
	Teneurin-4	Ten-m4[60]	Developing brain, spinal cord[60]
Gallus gallus	Teneurin-1[56, 61]		Developing CNS and eye[56, 61-63] limb buds[63]
	Teneurin-2[61, 64]		Developing brain and eye[61-63], AER of limb buds, tendon primordia, pharyngeal arches, heart, somites, neural tube, craniofacial mesenchyme[64]
	Teneurin-3[63]		Developing CNS, central retina, limb bud[63]
	Teneurin-4[63, 65]		Developing CNS, ZPA of limb buds, pharyngeal arches[65] heart, lung bud[63]
Rattus rattus	Teneurin-2	Neurestin[66]	Developing and adult CNS, somites[66]
Mus musculus	Teneurin-1	Ten-m1[67], odz1[68, 69], ten-m/odz1[70]	Developing and adult CNS[67, 69, 70], eye, smooth muscle cells in lungs, kidney glomeruli, adult testes[67]
	Teneurin-2	Ten-m2[67], odz1[71], odz2[68, 69], ten-m/odz2[70]	Developing and adult CNS[69, 70]
	Teneurin-3	Ten-m3[67], odz3[68, 69], ten-m/odz3[70]	Developing and adult brain[69-71], developing eye, spinal cord, notochord, craniofacial mesenchyme, tongue, dermis, saccule, developing limb, periosteum[68]
	Teneurin-4	Ten-m4[67], odz4[68, 69, 72], ten-m/odz4[70], DOC4[68, 71, 73]	Developing and adult brain[69-71, 73], developing eye, somites, spinal cord, trachea, nasal epithelium, saccule, joints, adipose tissue, tail bud and limbs[70]

adapted from [74]

The molecular functions of these domains in the ICD are still unknown.[75] The two isoforms found in *C.elegans* only differ in the length of the ICD.[55] The single transmembrane domain following the ICD contains around 23 hydrophobic amino acids. Directly adjacent to the transmembrane domain, there is a stretch of around 200 residues often, referred to as a linker region. This linker region is followed by eight tenascin-like EGF-like repeats. In the second and fifth EGF-like repeat, the free cysteines that are characteristic for teneurins can form lateral cross-links, resulting in the formation of dimers.[67, 76] The central part of the ECD contains 17 highly conserved cysteine residues. Structural prediction software shows NHL repeats in this region. NHL domains are known to form propeller-like twisted anti-parallel β-sheets.[77] This region of the protein is highly conserved, though its function remains yet to be elucidated. Closer to the C-terminus 26 YD repeats and a series of partial YD repeats are located. In eukaryotes, these repeats are unique to teneurins, but are also found in cell wall proteins of some prokaryotes.[50] The C-terminal half of the ECD has been found to form a globular domain, connected to the rod-like EGF-like repeats.[76] On the very C-terminal end of the teneurin protein a peptide sequence is found, which resembles the corticotrophin releasing factor (CRF) family of peptides. Synthetic peptides from this region called "teneurin C-terminal associated peptides" (TCAPs) reportedly influence neurite outgrowth and play a role in regulating behavioral responses to stress and anxiety.[78]

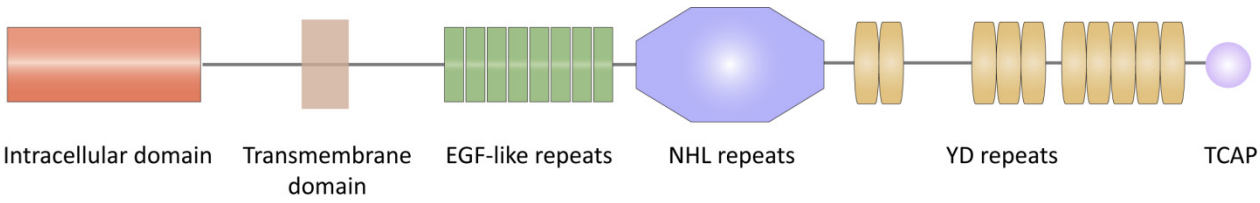


Figure 7: Domain architecture of vertebrate teneurins
All vertebrate teneurins share the same domain structure. They consist of an N-terminal intracellular domain (ICD), a single-span transmembrane domain and a large extracellular domain (ECD). The ECD consists of eight EGF-like repeats, a twisted NHL repeat domain and a YD repeat domain unique for eukaryotes. A peptide sequence (TCAP) related to the corticotrophin releasing factor is found at the C-terminal end.

2.3.2. Teneurin expression studies

Teneurins are mainly expressed in the developing central nervous system (CNS). Many reports have also shown teneurin expression in non-neuronal tissues such as the developing limbs and other sites of pattern formation (Table 1). In *C.elegans*, the two known isoforms show distinct expression patterns. The longer isoform Ten-1L is expressed in the somatic gonad, pharynx and in a subset of muscles and neurons.[55, 57, 79] The shorter isoform Ten-1S is found in hypodermal cells and in subsets of neurons.[55, 57, 79] In *Drosophila*, ten-a is expressed in the CNS, muscle attachment sites and the eye.[51, 59] Notably, besides the prominent expression in the neural and the optic system in the fly, ten-m is mainly expressed at sites contributing to the morphogenesis and patterning of the antennal discs, morphogenic furrow of the eye, the wing pouch and legs.[52, 53, 58] In zebra fish the expression patterns of only two members of the family were analyzed, namely teneurin-3 and -4. They have been found to be expressed in the spinal cord, somites and the developing brain.[60] The expression of teneurins has been most extensively studied in chicken. Here they are expressed in the nervous system and in sites of pattern formation. A well-described example for the latter is the limb bud. Here, teneurin-2 is exclusively expressed in the apical ectodermal ridge, whereas teneurin-3 is only expressed in the dorsal limb.[63, 64] Teneurin-1 is expressed in the dorsal ectoderm and the ventral mesenchyme.[63] Teneurin-4 is expressed in the distal and posterior mesenchyme.[63, 64] This complex pattern of expression of the different family members is also found in the nervous system. Different teneurins are expressed in distinct non-overlapping subpopulations of neurons in the developing visual system.[56, 61] In the developing optic tectum, the expression patterns of avian teneurin-1 and -2 are well-described.[61] More detailed studies of the distinct expression of different teneurins in different layers of the retina followed later.[62, 63, 80] In the mouse, most studies analyzed the expression patterns of teneurins in the developing nervous system, where they are also expressed in distinct subpopulations of neurons.[70] Prominent sites of expression in the brain are the pyramidal layer of the hippocampus, the dentate gyrus, the Purkinje cell layer in the cerebellum and the granular and molecular cell layer.[69]

2.3.3. Teneurin functions *in vitro* and *in vivo*

Many studies trying to shed light on teneurin functions and their mechanisms of action were performed *in vitro.* Neurite outgrowth was induced in neuroblastoma cells (Nb2a), transfected with full-length teneurin-2 and in explanted chick dorsal root ganglia when plated on teneurin-1 YD-repeats.[56, 61] Furthermore, the Nb2a cells transfected with teneurin-2 show enlarged growth cones.[61] Later on, this effect was shown to be dependent on the presence of the ICD domain, suggesting the need of a connection to the cytoskeleton.[81] The ICD of teneurin-1 was shown to interact with the cytoskeletal adapter protein CAP/ponsin, representing one possible link to the cytoskeleton.[80] Beside its possible adaptor function, the teneurin ICD seems to play an important role in mediating teneurin functions. Increasing evidence suggests, that teneurins are substrates for a process called "regulated intramembrane proteolysis" (RIP) releasing the ICD from the membrane. Overexpression of soluble teneurin-2 ICD constructs in HT1080 cells leads to an accumulation of the ICD in the nucleus and a co-localization with PML bodies.[82] Investigations in the developing chicken brain showed nuclear stainings for the teneurin-1 ICD.[62] Furthermore, stainings of *C.elegans* with an antibody against an N-terminal peptide showed nuclear stainings of embryos and adult gut cells.[55] RIP requires two steps. In a first step, the protein substrate is cleaved by a protease near the transmembrane domain at the outside of the cell, releasing the extracellular protein part into the extracellular milieu. In a second step, a protease cleaves the substrate in the transmembrane domain or closeby inside the cell, releasing the intracellular part from the membrane into the cell.[83] Analysis of the teneurin sequences revealed several conserved protease cleavage sites.[50] Most of the teneurins possess a conserved furin-like cleavage site between the transmembrane domain and the EGF-like repeats. This cleavage site was shown to be functional *in vitro*, by using a recombinant avian teneurin-2 fusion protein.[61, 64] This fulfills the first step required for RIP. The protease responsible for releasing the teneurin ICD from the membrane is not yet known. Possible candidate proteases are signal peptide peptidase (SPP), SPP-like proteases and site-2 protease, which have all been shown to cleave type II transmembrane proteins.[84] Nevertheless, the nuclear function of the ICD is still unknown. A yeast-two hybrid screen for identifying interaction partners of the teneurin-1 ICD revealed MBD1, a nuclear methyl CpG binding protein as a novel binding partner.[80] MBD1 is known for its function in transcriptional repression.[85] In a different study, the expression of the teneurin-2 ICD

inhibited the transcriptional activity of ZIC1.[82] The experiments suggest that the teneurin ICD may be involved in transcriptional activity. Some studies suggest that teneurins interact homophilically.[81, 82] However, different studies also show interactions of teneurins with ECM proteins and basement membrane proteins.[61, 64, 79, 86] Furthermore, a recent study identified latrophilin-1 as a transsynaptical binding partner of teneurin-2.[87]

Accumulating evidence is beginning to shed light on the *in vivo* function of teneurins in different species. A knock-out of the single teneurin gene in *C.elegans* resulted in a pleiotrophic phenotype. The mutants showed an increased embryonic lethality, gonad disorganization, defects in distal tip cell migration, nerve cord defasciculation, defects in axonal path finding and basement membrane integrity.[55, 86] Milder versions of these phenotypes were also observed in a different study, in which a truncated form of the *C.elegans* teneurin that ended just after the EGF repeats was used.[57] This shows the importance of the domains, located C-terminally of the EGF-like repeats in mediating teneurin functions. In *Drosophila,* mutant flies for ten-m showed defects in retinal development and motor neuron guidance.[54, 88] In vertebrates, cell bodies of neurons transfected with a teneurin-3-GFP construct *in utero* aggregate and are delayed in the migration to their target region.[89] However, teneurin-3-GFP overexpression led to an increased neurite outgrowth, as observed for teneurin-2 *in vitro*.[50, 89] Teneurin-3 knock-out mice have a defect in axon targeting of the ipsilateral projections in the visual system, resulting in an impairment of binocular vision.[90]

2.3.4. Teneurin-1 as a candidate gene for X-linked mental retardation

Patients suffering from mental retardation (MR) show a complex phenotype, with significant limitations in their intellectual capacity and adaptive behavior. The onset of the disease is before the patients reach 18 years of age without showing a degenerative phenotype.[91] The most severe forms of MR, in which patients have an IQ lower than 50 is very likely to be caused by specific genetic contributions.[92] Among the monogenic forms of MR, X-linked mental retardation (XLMR) is the most frequent one. To date, mutations in more than 90 genes located on the X-chromosome have been implicated in causing XLMR.[93] Gene products of the identified XLMR genes perform different biological functions, such as transcriptional regulation, RNA processing, cell cycle regulation and signal transduction.[94] More than 20 of these proteins produced by these genes are localized to synapses, implicating a role in synaptogenesis, synaptic activity and plasticity as underlying causes of XLMR.[95] *Odz1*, the gene coding for teneurin-1, resides in the gene locus Xq25. This locus has a very low gene density and is implicated in causing XLMR in several families. Three XLMR-causing genes were identified in Xq25. These genes are called *UPF3B, ZDHHC9* and *GRIA3.* The latter of these genes is located very close to the teneurin-1 locus.[96-98] Given the established functions of teneurins in the developing brain as axon guidance cues and in target recognition, teneurin-1 may be a candidate gene for causing XLMR. Furthermore, the known interaction partner of the teneurin-1 ICD, MBD1, is also implicated in causing MR.[99]

3. Aim of the thesis

The teneurin protein family is a unique family of type II transmembrane proteins conserved from worms and flies to vertebrates. Studies of different members of this family in invertebrates and vertebrates suggest an important role during embryonic development, especially in the developing brain. Despite of their high conservation in the domain structure and amino acid sequence, a systematic study of the origin of teneurin in evolution and a detailed comparison of the conserved domains between species was still missing. The result of such a study is presented in the first part of the result section (chapter 4.1.).

In vertebrates, the existing four paralogs, called teneurin-1 to -4 are expressed in the developing brain in distinct non overlapping patterns. However, there is still very little known about which factors are regulating teneurin expression and hence their expression patterns, which seem to be a key factor for teneurin-mediated functions. Furthermore, teneurin-1 has been implicated to represent a potential target gene in XLMR. In order to provide a basis for further deletion and mutation studies of teneurin-1 in XLMR patients, the 5' UTR and gene promoter elements of teneurin-1 was characterized. Furthermore, the transcription factor binding to and regulating a newly discovered promoter element was analyzed in the second part of this thesis (chapter 4.2.).

Several studies had shown that teneurins interact homophilically. However, which of the teneurin domains is responsible for this interaction remained yet to be elucidated. The aim of the third part of this thesis was to investigate the teneurin domains contributing to the homophilic interaction and investigate the functional consequences of such interactions for neurons (chapter 4.3.).

4. Results

4.1. Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry

Richard P. Tucker, Jan Beckmann, Nathaniel T. Leachman, Jonas Schöler
and Ruth Chiquet-Ehrismann

Mol Biol Evol *(2012) 29 (3): 1019-1029*

First published online: October 31, 2011

My contribution to this paper:

For this study, I performed database analysis to find known transcript and EST sequences for the various species analyzed. Furthermore, I did initial sequence alignments and gave input to the manuscript.

Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry

Richard P. Tucker,*[,1] Jan Beckmann,[2,3] Nathaniel T. Leachman,[1] Jonas Schöler,[2,3] and Ruth Chiquet-Ehrismann[2,3]

[1]Department of Cell Biology and Human Anatomy, University of California at Davis

[2]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Basel, Switzerland

[3]Faculty of Science, University of Basel, Basel, Switzerland

***Corresponding author:** E-mail: rptucker@ucdavis.edu.

Associate editor: Billie Swalla

Abstract

Teneurins are type II transmembrane proteins expressed during pattern formation and neurogenesis with an intracellular domain that can be transported to the nucleus and an extracellular domain that can be shed into the extracellular milieu. In *Drosophila melanogaster*, *Caenorhabditis elegans*, and mouse the knockdown or knockout of teneurin expression can lead to abnormal patterning, defasciculation, and abnormal pathfinding of neurites, and the disruption of basement membranes. Here, we have identified and analyzed teneurins from a broad range of metazoan genomes for nuclear localization sequences, protein interaction domains, and furin cleavage sites and have cloned and sequenced the intracellular domains of human and avian teneurins to analyze alternative splicing. The basic organization of teneurins is highly conserved in Bilateria: all teneurins have epidermal growth factor (EGF) repeats, a cysteine-rich domain, and a large region identical in organization to the carboxy-half of prokaryotic YD-repeat proteins. Teneurins were not found in the genomes of sponges, cnidarians, or placozoa, but the choanoflagellate *Monosiga brevicollis* has a gene encoding a predicted teneurin with a transmembrane domain, EGF repeats, a cysteine-rich domain, and a region homologous to YD-repeat proteins. Further examination revealed that most of the extracellular domain of the *M. brevicollis* teneurin is encoded on a single huge 6,829-bp exon and that the cysteine-rich domain is similar to sequences found in an enzyme expressed by the diatom *Phaeodactylum tricornutum*. This leads us to suggest that teneurins are complex hybrid fusion proteins that evolved in a choanoflagellate via horizontal gene transfer from both a prokaryotic gene and a diatom or algal gene, perhaps to improve the capacity of the choanoflagellate to bind to its prokaryotic prey. As choanoflagellates are considered to be the closest living relatives of animals, the expression of a primitive teneurin by an ancestral choanoflagellate may have facilitated the evolution of multicellularity and complex histogenesis in metazoa.

Key words: Teneurin, Odz, Ten-m, evolution, horizontal gene transfer, choanoflagellate, *Monosiga brevicollis*.

Introduction

Teneurins are phylogenetically conserved transmembrane proteins (see reviews by Tucker and Chiquet-Ehrismann 2006; Young and Leamey 2009). The name "teneurin" honors their discovery in *Drosophila melanogaster* by combining the names of the two dipteran teneurin homologs, Ten-a (Baumgartner and Chiquet-Ehrismann 1993) and Ten-m (Baumgartner et al. 1994, also referred to as Odz [Levine et al. 1994]), with neurons, which are common sites of expression (e.g., Minet et al. 1999). In *D. melanogaster*, chicken, and mouse the teneurin homologs have the following conserved features: 1) teneurins are type II transmembrane proteins with an N-terminal intracellular domain (ICD) and a large extracellular domain (ECD); 2) teneurins have eight epidermal growth factor (EGF) repeats; 3) the third cysteine residue in the second and fifth EGF repeat is replaced with a tyrosine or phenylalanine residue, which results in the potential for teneurins to dimerize side-by-side through disulfide bonds (Oohashi et al. 1999); and 4) the C-terminal two-thirds of teneurins is similar to the YD-repeat proteins of prokaryotes, with characteristic NHL (from NCL-1, HT2A, and Lin-41) repeats, tyrosine and aspartate-rich YD repeats, and a region similar to the core-associated domain of retrotransposon hot spot (RHS) proteins. In addition, many teneurins can be proteolytically processed, freeing the ICD for transport to the nucleus (Bagutti et al. 2003; Nunes et al. 2005; Kenzelmann et al. 2008) and/or releasing the ECD for incorporation in the extracellular matrix (ECM; Rubin et al. 1999; Trzebiatowska et al. 2008). An additional cleavage site near the C-terminus can lead to the creation of a neuropeptide (reviewed by Lovejoy et al. 2009). Proline-rich Src homology 3 (SH3)–binding domains have been identified in the ICD of teneurins cloned from chordates and ecdysozoans, and ICD-interacting partners have been characterized that may mediate associations between teneurins and the cytoskeleton and methylated DNA (Nunes et al. 2005). Mutation analysis and RNAi-mediated knockdown of teneurin expression in *D. melanogaster* and *Caenorhabditis elegans* reveal fundamental roles for teneurins in

pattern formation (Baumgartner et al. 1994; Rakovitsky et al. 2007), axonal fasciculation (Drabikowski et al. 2005), and the integrity of basement membranes (Trzebiatowska et al. 2008). In chordates, teneurins are best studied in mouse and chicken, where they are predominantly expressed in the developing nervous system in area-specific patterns mediated in part by EMX2 (Li et al. 2006; Beckmann et al. 2011). Knockout of the gene encoding mouse teneurin-3 by homologous recombination results in abnormal pathfinding in the visual system and a loss of binocular vision (Leamey et al. 2007).

In order to identify novel features and learn more about the potential evolutionary origin of teneurins, we searched for and compared sequences encoding teneurin-like proteins in opisthokont genomes and collections of expressed sequence tags (ESTs). We also cloned and sequenced cDNAs encoding the ICD of teneurins from human and chicken to study alternative splicing. By aligning and analyzing proteins for predicted nuclear localization sequences (NLSs), SH3-binding domains, and furin-type proteolytic cleavage sites, we have refined our knowledge of conserved teneurin structure and function. In addition, we identified a gene encoding a teneurin in the choanoflagellate *Monosiga brevicollis*, which suggests that teneurins may have played a role in the early evolution of metazoan tissues.

Materials and Methods

Sequence Analysis

Novel teneurin sequences were identified by sequence homology using tBLASTn (http://blast.ncbi.nlm.nih.gov/) and by domain architecture using Pfam (http://pfam.sanger.ac.uk/) with "view a family," SMART (http://smart.embl-heidelberg.de/smart/) with "architecture queries," and Superfamily (http://supfam.org/SUPERFAMILY/) with "domain combinations." Boundaries of regions, domains, and repeats were determined using Pfam for EGF repeats, NHL repeats, RHS repeats (related to YD repeats), RHS protein, Ten_N domains, and PfamB PB025792 (the region between the transmembrane domain and the EGF repeats, which was identified as a phylogenetically conserved region by Pfam), and SMART for transmembrane domains and EGF repeats. Alignments and phylogenetic relationships were determined using ClustalW (http://www.genome.jp/tools/clustalw/) and the settings "pair alignment slow/accurate" (gap open penalty 10, gap extension penalty 0.1). Importin α/β pathway NLSs were identified using NLS Mapper (http://nls-mapper.iab.keio.ac.jp), and furin cleavage sites were predicted with ProP (http://www.cbs.dtu.dk/services/ProP/). SH3-binding domains were identified by hand from consensus sequences described by others (Kay et al. 2000; Mayer 2001; Kowanetz et al. 2003).

Reverse Transcriptase PCR and Sequencing

Human adult brain cDNA was generated out of total human brain RNA (AMS Biotechnology, Oxon, UK) using Superscript III (Invitrogen, Carlsbad, CA) polymerase and random hexamer primers (Invitrogen). Sequences corresponding to the ICDs of teneurins-1 through -4 were amplified with PfuTurbo polymerase (Stratagene/Agilent Technologies, Santa Clara, CA) using specific primers (teneurin-1: 5′-ACTAGCGGCCGCACCATGGAGCAAACTGACTGC-3′/5′-ACTACTCGAG GCAGCACCTGTAAGGTTTG-3′; teneurin-2: 5′-ACTAGCGGCCGCACCATGGATGTAAAGGACCGG-3′/5′-ACTACTCGAGGCAGTATTTGGAGGGCTTC-3′; teneurin-3: 5′-ACTAGCGGCCGCACCATGGATGTG AAAGAACGC-3′/5′-ACTACTCGAGACAGTACTTTGAAGACTTC-3′; teneurin-4: 5′-ACTAGCGGCCGCACCATGGAC GTGAAGGAGAGG-3′/5′-ACTACTCGAGACAGTACTTGG AGGGCTTC-3′) including restriction sites NotI and XhoI. Amplified products were separated on a 0.8% agarose gel, and fragments were excised, gel extracted, and cloned into pcDNA3. Positive clones were sequenced using forward primer T7 and reverse primer Sp6.

The sequence of avian teneurin-3 including alternatively spliced variants was assembled from overlapping fragments cloned by polymerase chain reaction (PCR). cDNA was prepared from total RNA extracted from embryonic day 16 chicken cerebellum using the RNeasy Mini kit (Qiagen, Germantown, MD). PCR was performed with the Platinum Pfx DNA Polymerase System (Invitrogen). Five sets of primers were used to divide avian teneurin-3 into five segments. Segment 1 used primer pair 5′-ATGGATGTGAAAGAACGTCG-3′/5′-CACGTGGAGGGTAAACGATAA-3′; segment 2 used primer pair 5′-ACTGTGAAGAAGCGGATTGC-3′/5′-GACCGCCAAAAGTCACTAGA-3′; segment 3 used primer pair 5′-TGATGGGACCATGATCAGAA-3′/5′-ACCAGACGGCAGACATGAAC-3′; segment 4 used primer pair 5′-AGCGAGGGACGACTAGTGAA-3′/5′-GGAGAAAGGATAGAGTGAAA-3′; and segment 5 used primer pair 5′-AGGCTGTGGACAGAAGGAGA-3′/5′-GGTCCTCTACTTGGATGACT-3′. Each segment was TOPO cloned into the pCR-II vector (Invitrogen) for sequencing.

Results

Overview: Identification of Teneurins and Analysis of Teneurins from *Homo sapiens*

Genes encoding teneurin-like proteins and predicted proteins with the characteristic domain organization of known teneurins were identified by sequence similarity (e.g., tBLASTn) and by the presence of combinations of domains (e.g., predicted proteins with both EGF repeats and NHL repeats using Pfam or SMART; for details, see Materials and Methods). Teneurins identified in this way from chordates are summarized in supplementary table S1, Supplementary Material online; teneurins from nonchordates are summarized in supplementary table S2, Supplementary Material online.

To illustrate the features of these teneurins identified through proteomic analysis, the four teneurins from *H. sapiens* are shown in figure 1. The variant of teneurin-1 shown in figure 1 is a type II transmembrane protein with a 317aa N-terminal ICD, a 23aa transmembrane domain, and a 2385aa C-terminal ECD. Within the ICD, there are two predicted importin α/β pathway NLSs. The first NLS (aa11-40) has an NLS Mapper score of 4.8, and the second NLS (aa60-69) has an NLS Mapper score of 6.0. Higher

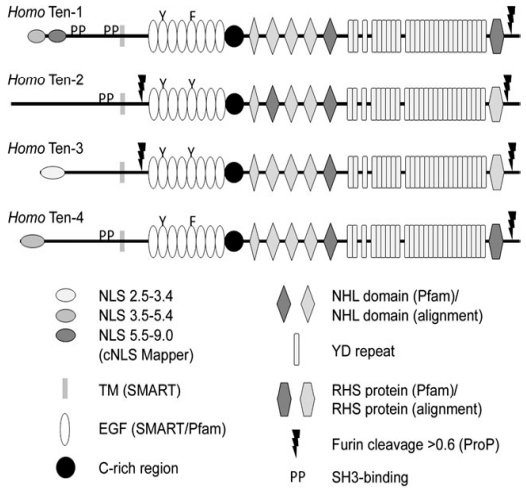


Fig. 1. The human teneurins. Human teneurins share a basic domain organization but they differ in the presence of NLSs, SH3-binding motifs, and furin cleavage sites. Following processing the ICD of teneurin-1, for example, is predicted to be located in the nucleus but not the ICD of teneurin-2. Similarly, the ECD of teneurins-2 and -3 are predicted to be shed into the ECM but not the ECD of teneurins-1 and -4. Accession numbers, the positions of domains as well as the NLS and ProP scores of each teneurin are found in supplementary table S3, Supplementary Material online.

scores represent a greater likelihood of nuclear localization (Kosugi et al. 2009) and are indicated on the figure with progressively darker ovals. Two proline-rich SH3-binding domains (indicated by "PP" on fig. 1) are also found in the ICD of human teneurin-1. The first (aa193-199; RPLPPPP) is consistent with the consensus sequence for Class I SH3 ligands (+xφPxφP); the second (aa292-297; PRPLPR) is consistent with the atypical SH3-binding motif (PxxxPR) of cbl proteins (Kowanetz et al. 2003). In the ECD of human teneurin-1, there are eight EGF repeats (aa531-796). The third cysteine residue of the second EGF repeat has been replaced with a tyrosine residue, and the third cysteine residue of the fifth EGF repeat has been replaced with a phenylalanine (indicated by the "Y" and "F" in fig. 1). This substitution results in the potential for dimerization of teneurin monomers through disulfide bonds between cysteines that lack an intramodular partner (Oohashi et al. 1999). A cysteine-rich domain is found adjacent to the eighth EGF repeat (aa797-836). The carboxy two-thirds of human teneurin-1 shares the same domain organization as the YD-repeat proteins of some prokaryotes (e.g., *Myxococcus xanthus*, where it is required for gliding motility [Youderian and Hartzell 2007]): five NHL domains, YD-repeats (similar to "RHS repeats"), and a region near the C-terminal tail identified by Pfam as RHS protein (similar to "RHS-associated core domain"). The NHL domains of human teneurin-1 were identified by Pfam (dark gray, fig. 1) or by alignment using ClustalW (light gray, fig. 1). Similarly, the RHS protein domain identified by Pfam in human teneurin-1 is indicated in dark gray, and those identified in other teneurins by alignment are indicated by lighter shades. Finally, human teneurin-1 has a single predicted furin cleavage site with a ProP score at or above 0.55 (threshold = 0.50) at aa2618 (LNGRTRR/FA). This would create a 107aa C-terminal peptide similar to teneurin C-terminal-associated peptide-1 (Trubiani et al. 2007).

There are four teneurin genes in *H. sapiens*. The basic organization of teneurins-1, -2, -3, and -4 is the same, most notably in the ECD: each teneurin has eight EGF repeats with aromatic residues substituting for cysteines in the second and fifth repeat, each has a cysteine-rich domain and a C-terminal region similar to the YD-repeat proteins of prokaryotes, and each has a predicted furin cleavage site near the C-terminus (fig. 1; details can be found in supplementary table S3, Supplementary Material online). One difference is the presence of a second predicted cleavage site between the transmembrane domain and EGF repeats of teneurins-2 and -3 that would permit shedding of the ECD into the ECM; these cleavage sites are not found in teneurins-1 and -4. Additional differences are seen in the ICD. Teneurins-1, -2, and -4 have proline-rich motifs that match consensus SH3-binding domains but teneurin-3 does not. However, the proline-rich sequence PPTRPLPR is found in the ICD of human teneurin-3, which resembles, but does not exactly match, known SH3-binding motifs. Teneurin-2 does not have a predicted NLS, and the NLS of teneurin-3 has a lower NLS Mapper score than the NLSs of teneurin-1 and teneurin-4.

Representative teneurins from major taxonomic groups were analyzed in this manner and are described below.

Identification and Analysis of Chordate Teneurins

The teneurins of mouse (*Mus musculus*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), and the protochordates *Ciona intestinalis* and *Branchiostoma floridae* were identified and analyzed. There are four teneurins in *M. musculus* and *G. gallus* and they share many of the features described above for human teneurins. The few differences include the absence of a predicted NLS in murine teneurin-4, and the observation that chicken teneurin-2 has a predicted NLS, albeit a weak one (fig. 2*A* and *B*; supplementary table S3, Supplementary Material online). In *D. rerio*, there are five teneurins. ClustalW alignment and basic phylogenomic analysis identify two teneurin-2 paralogs that we have named teneurin-2A and teneurin-2B (figs. 2*C* and 3; supplementary table S3, Supplementary Material online). The predicted sequences of teneurin-1, teneurin-2B, and teneurin-4 appear to be complete, but the predicted N-termini of teneurin-2A and teneurin-3 were completed by translating potential open reading frames and aligning them with known teneurin sequences and by piecing together ESTs (supplementary table S3, Supplementary Material online; FASTA files can be found in supplementary table S4, Supplementary Material online). As with the chicken teneurins, the basic features of the zebrafish teneurins are conserved. Differences include the absence of a potential furin cleavage site near the C-terminus of teneurin-1, an additional potential furin cleavage site between the NHL repeats and YD repeats of teneurin-2A and the absence of predicted NLSs

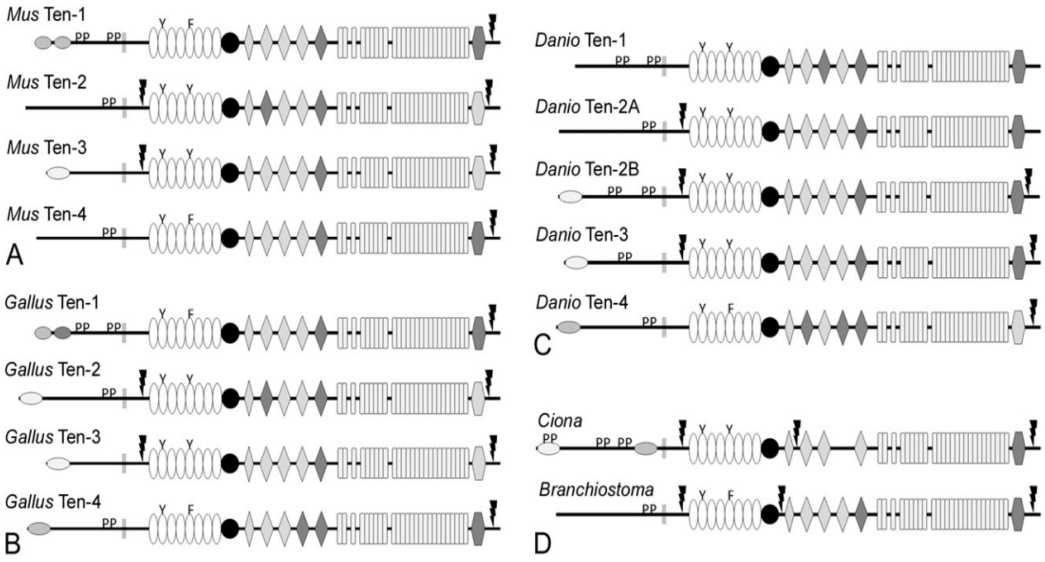


FIG. 2. The domain organization and possible relationships of chordate teneurins. There are four teneurins in mouse (*Mus musculus*) and chicken (*Gallus gallus*). Basic features are highly conserved between mouse (*A*), chicken (*B*), and human teneurins (fig. 1). There are five teneurin genes in the zebrafish, *Danio rerio*, including two teneurin-2 paralogs. The ICD of the predicted teneurin-1 does not have an NLS, and an additional furin cleavage site is found between the NHL domains and the YD repeats (*C*). The genomes of the tunicate *Ciona intestinalis* and the cephalochordate *Branchiostoma floridae* each contain a single teneurin gene (*D*).

in the ICDs of teneurin-1 and teneurin-2A. In addition, the ICD of *D. rerio* teneurin-3 has a predicted proline-rich SH3-binding domain, unlike the ICDs of teneurin-3 in chicken and man.

To determine if the duplication of teneurin-2 is a common feature in bony fish, the teneurins of the stickleback *Gasterosteus aculeatus* were also identified (supplementary table S1, Supplementary Material online) and aligned with the teneurins of other chordates. Like *D. rerio*, *G. aculeatus* has five teneurin genes. However, there are two teneurin-3 paralogs (teneurin-3A and teneurin-3B) and only one teneurin-2 (fig. 3). The teneurin-1 of *G. aculeatus* has a potential furin cleavage site near the C-terminus (aa2642; ProP score = 0.74), indicating that the absence of this site in *D. rerio* may not be typical of teneurin-1 in actinopterygians. The *G. aculeatus* teneurin-1 also has a predicted NLS in the ICD (aa11-41, NLS Mapper score = 4.0).

The genomes of *C. intestinalis* and *B. floridae* each encode a single teneurin (fig. 2*D*, supplementary table S3, Supplementary Material online). The basic organization of these teneurins resembles those of the craniate teneurins. The ICD of the predicted teneurin from *C. intestinalis* has two NLSs: one near the N-terminus and the other near the transmembrane domain. A predicted NLS near the transmembrane domain is commonly observed in the ICD of teneurins from protostomes (see below). The ICDs from both of the protochordate teneurins have predicted SH3-binding domains, but the ICD from *B. floridae* does not have an NLS that is recognized by NLS Mapper. Both of the predicted protochordate teneurins have potential furin cleavage sites that would shed the ECD and process the C-terminus like those of teneurin-2 and teneurin-3 in fish, birds, and man, but like *D. rerio* teneurin-2A (and unlike other craniate teneurins examined) they also have a third predicted furin cleavage site near the center of the ECD. A second teneurin-like sequence is found in the *B. floridae* genome when two adjacent predicted proteins (XP_0025 92160 and XP_002592161) are combined. However, the C-terminal two-thirds of the second predicted protein also

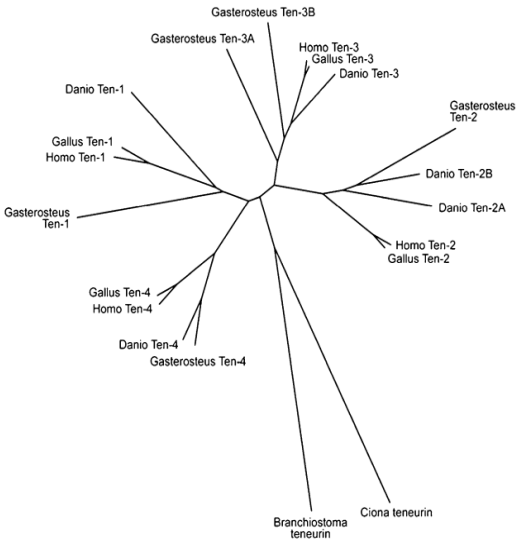


FIG. 3. An unrooted phylogenetic tree based on ClustalW alignment of real and predicted teneurin amino acid sequences suggests that teneurins-1 and -4 share a common ancestor, as do teneurins-2 and -3. The stickleback *Gasterosteus aculeatus* has five teneurins, but unlike *Danio rerio* it has retained two teneurin-3 paralogs.

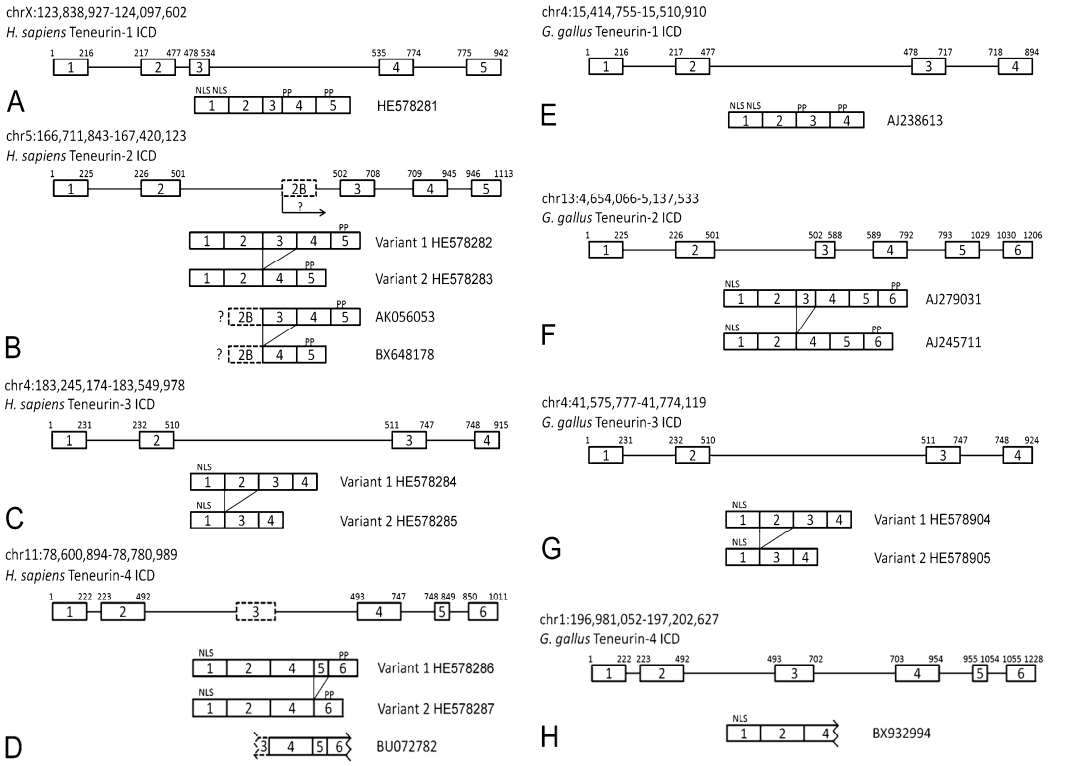


FIG. 4. Comparison of alternative splicing in the ICD of human and chicken teneurins. The human teneurin-1 ICD is encoded on five exons. No splice variants were found using RT-PCR and primers corresponding to sequences in exon 1 and exon 5 (*A*). In contrast, RT-PCR reveals two variants amplified with primers based on sequences in the first and fifth exon of human teneurin-2. Variant 1 is encoded on 5 exons, whereas Variant 2 is composed of four. ESTs suggest an additional way to generate diversity in the ICD of human teneurin-2: An exon found between the second and third exon (exon 2B) is associated with long and short variants as well. This may represent an alternative start site for teneurin-2 (*B*). Human teneurin-3 has an ICD encoded on four exons, and exon 2 can be spliced out to generate a second variant (*C*). RT-PCR with primers based on sequences in exon 1 and exon 6 reveal long and short variants of human teneurin-4. Variant 1 is encoded on five exons, whereas Variant 2 is encoded on four. Use of an additional exon (exon 3) that does not contain a start codon is found in an EST (*D*). The ICDs of teneurins-1 and -2 from the chicken were described previously. As in human, there is a single teneurin-1 isoform, and there are long and short variants of teneurin-2. There is no evidence from chicken of an alternative start site in teneurin-2 (*E*, *F*). PCR using cDNA from embryonic chicken cerebellum and primers based on sequences in exons 1 and 4 reveal identical variants in chicken and human teneurin-3 (*G*). Sequence from a single EST from the chicken is consistent with the organization of the ICD of human teneurin-4 (*H*).

has lysozyme and keratin-related sequences, and some teneurin and lysozyme-like sequences are encoded on the same large predicted exon, which leads us to suggest that this is a pseudogene. This is supported by the total absence of ESTs.

Alternatively Spliced Variants

Previously we showed that there are a number of isoforms of avian teneurin-2 and that these variants are derived from alternative splicing of regions of transcripts encoding both the ICD and the ECD (Tucker et al. 2001). Here, we examined ESTs and used RT-PCR to determine if the ICD variants are specific for teneurin-2 and if they are conserved in mammals and birds. A single PCR product is amplified from adult human brain-derived cDNA using primers corresponding to the 3′ end of the first exon of the human teneurin-1 gene and the 5′ end of the fifth exon, which encodes the transmembrane domain (fig. 4*A*). When the same strategy is applied using primers based on human teneurin-2 sequences, two variants are observed: Variant 1 is encoded by all five previously identified exons and Variant 2 lacks the third exon (fig. 4*B*). These variants are supported by EST data, which also reveal the use of a sixth exon found between exon 2 and exon 3. EST sequences containing this alternative exon, which we have named exon 2B, do not contain sequences corresponding to either exon 1 or exon 2. As a putative start codon is found in exon 2B, this exon may be used as an alternative start site for teneurin-2 transcripts (and therefore would not have been amplified using our flanking primer pairs). The ICD of teneurin-3 is encoded on four exons and like teneurin-2 there are two ICD splicing variants: Variant 1 uses all four exons, whereas Variant 2 is encoded on exons 1, 3, and 4 (fig. 4*C*). Finally, RT-PCR reveals two alternatively spliced variants of the human teneurin-4 ICD. The larger is encoded by five exons, and a smaller by four exons. Interestingly, an EST (BU72782) shows the potential use of an additional exon that was not amplified by our primer pair (fig. 4*D*).

ESTs demonstrate that some of the ICD variants we observed in human teneurin-1 and teneurin-4 are conserved

Tucker et al. · doi:10.1093/molbev/msr271 MBE

in *G. gallus*, just as our previous work with avian teneurin-2 showed the presence of two ICD variants (fig. 4*E*, *F*, and *H*). To study the alternative splicing of teneurin-3, we used RT-PCR to amplify products corresponding to the ICD and cDNA derived from embryonic chicken brain. Just as in human, the avian teneurin-3 ICD is encoded on four exons. A large variant contains sequences corresponding to all four exons, and exon 2 is spliced from a smaller variant (fig. 4*G*). Note that we could not identify an exon homologous to exon 2B of human teneurin-2 in the chicken genomic sequence, but there is a homologous potential exon 3 in chicken teneurin-4 DNA.

There is also evidence of teneurin variants derived from alternative splicing in the region encoding the ECD. For example, a short (8aa) stretch of amino acids can be present between the seventh and eighth EGF repeats of teneurin-2 from the chicken, and at least one variant of avian teneurin-2 is truncated between the seventh and eighth EGF repeats, resulting in an isoform lacking the cysteine-rich domain and the region homologous to the YD repeat proteins of prokaryotes (Tucker et al. 2001). Alternative splicing that results in additional sequence between the seventh and eighth EGF repeats may be common in teneurins, as similar variants are found in mRNA sequences in mouse teneurin-3 (NP_035987) and teneurin-4 (BAE28005). However, there is no evidence supporting grossly truncated isoforms of teneurins in other species.

Identification and Analysis of Teneurins from an Echinoderm and Protostomes

The same methods used to identify teneurins in chordates were applied to other metazoan sequences. In addition to the known teneurins of *D. melanogaster* and *C. elegans*, predicted complete or partial teneurins were found in the purple sea urchin (*Strongylocentrotus purpuratus*), a mollusk (*Lottia gigantean*), an annelid (*Capitella teleta*), a trematode (*Schistosoma mansoni*), and a wide variety of nematodes and arthropods (supplementary table S2, Supplementary Material online). Interestingly, no teneurin-like sequences were identified in the genomes or ESTs from cnidarians, ctenophores, sponges, or *Trichoplax adhaerens*.

The single teneurin from *S. purpuratus* is remarkable for a few features not seen in teneurins from chordates: 1) it has only six EGF repeats and only the second EGF repeat has an aromatic residue substituting for a cysteine residue; 2) it lacks a predicted furin cleavage site near the C-terminus; and 3) it has two predicted furin cleavage sites between the transmembrane domain and the EGF repeats (fig. 5*A*). Like the teneurins from protochordates, it has a predicted furin cleavage site near the center of the ECD.

The teneurins from *C. elegans* and *D. melanogaster* are well known, and the sequences that were analyzed here came from cDNAs. There are two teneurins from *C. elegans*, Ten-1L and Ten-1S; they are encoded by the same gene, but two different promoters regulate the expression of "long" and "short" transcripts (Drabikowski et al. 2005). Ten-1L is illustrated in figure 4*A*; Ten-1S would be identical except the ICD is much smaller (the approximate location of

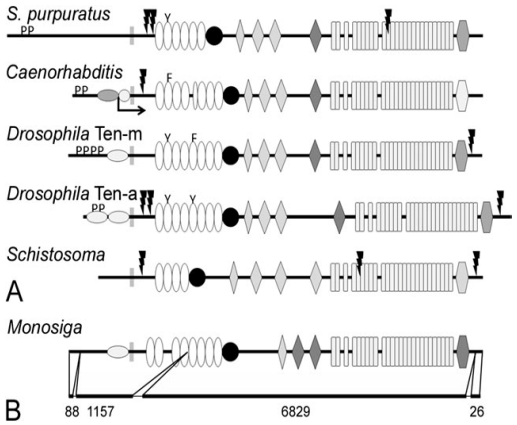


FIG. 5. The domain organization of nonchordate teneurins. The ECDs of teneurins from the purple sea urchin (*Strongylocentrotus purpuratus*), *Caenorhabditis elegans*, *Drosophila melanogaster*, and the trematode *Schistosoma mansoni* have the same basic organization as chordate teneurins, but there is variation in the number of EGF repeats. Aromatic residues substitute for cysteines in at least one EGF repeats in all the species examined except the fluke, which suggests that the teneurin from *S. mansoni* does not dimerize (*A*). The genome of the choanoflagellate *Monosiga brevicollis* encodes a protein with a domain organization that is identical to metazoan teneurins. The predicted protein does not have furin cleavage sites or SH3-binding domains, and its EGF repeats contain a full complement of cysteines. The predicted *M. brevicollis* teneurin is encoded on just four exons, and most of the ECD is encoded on a single mega-exon of 6829 bp (*B*).

the N-terminus of Ten-1S is indicated by the crooked arrow in fig. 5*A*). Ten-1L and the two *D. melanogaster* teneurins, Ten-m and Ten-a, have putative SH3-binding domains and one or more NLS. Unlike the NLSs of most chordate teneurins, the NLSs from the ecdysozoans tend to be found near the transmembrane domain and not at the N-terminus. The ECDs of these teneurins are similar to those found in chordate teneurins: note that both Ten-m and Ten-a (but not Ten-1L) have potential furin cleavage sites near the C-terminus, and both Ten-1L and Ten-a have potential furin cleavage sites that could shed the ECD into the ECM. The fifth EGF repeat of *C. elegans* Ten-1L is incomplete; the part of this repeat encoding both the second and third cysteine residues in other teneurins is missing. Also, the part of the ECD near the C-terminus that is predicted by Pfam to be homologous to the RHS core-associated protein domain is more unlike this domain in other teneurins, though some identity could be found by alignment.

Two teneurin genes that encode predicted proteins that align with either *D. melanogaster's* Ten-a or Ten-m were identified in the genomes of a number of insect species, including *Apis mellifera* (honey bee), *Tribolium castaneum* (flour beetle), and the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus* (supplementary table S2, Supplementary Material online). However, single teneurin genes were found in the genomes of the branchiopod crustacean *Daphnia pulex* and the arachnid *Ixodes scapularis* (deer tick). This suggests that the duplication of teneurins in

Table 1. Alignment of Representative Sequences with the Cysteine-Rich Domain Core Sequence of *Monosiga brevicollis* Teneurin.

Species (common name)	Core Sequence	% Id	% Sim
Monosiga brevicollis (choanoflagellate)	CNDGIDNDNDRVTDCNDADCCSS		
Teneurin			
Phaeodactylum tricornutum (diatom)	CNDGIDNDNDGLFDCEDPDCAND		
Endo-1,3-beta-glucosidase	********** : **:*.**...	65%	91%
Schistosoma mansoni (fluke)	CDDGIDNDHDDLVDCLDPDCCTS		
Teneurin	*:******:* :.** *.***:*	65%	91%
Caenorhabditis elegans (roundworm)	CDDGLDNDSDGLIDCDDPECCSS		
Ten-a	*:*.****.* :***.*::***	61%	91%
Homo sapiens (human)	CGDNLDNDGDGLTDCVDPDCCQQ		
Teneurin-1	*.*.:***.* :*** *.***..	57%	91%
Volvox carteri (volvox)	CDDGIDNDCDGLVDMDDPDCNTS		
Gametolysin	*:****** * :.* :*.** :*	57%	83%
Strongylocentrotus purpuratus (purple sea urchin)	CTDEVDNDGDSLIDCEDPDCCLS		
Teneurin	*.* :***.* : **:*.*** *	57%	83%
Pfam: Cu-binding_MopE	C.DGVDNNCDGQVD		
PF11617*	* **:**: * .*	54%	77%

NOTE.—*The Cu^{++}-binding consensus domain of MopE and related proteins.

the protostome lineage is limited to insects and is not a feature of all arthropods.

The trematode *S. mansoni* has a single predicted teneurin (fig. 5*A*; supplementary table S3, Supplementary Material online) that has a number of distinctive features. Its ICD is relatively short, and it does not contain predicted SH3-binding motifs or an NLS. Like many teneurins it has a putative furin cleavage site between the transmembrane domain and the EGF repeats, but it also has a second predicted furin cleavage site amidst the YD repeats. Finally, the fluke teneurin has only four EGF repeats, and all four EGF repeats have a full complement of cysteine residues, so unlike other teneurins studied to date it probably fails to dimerize.

A Teneurin Is Encoded in the Genome of the Choanoflagellate *M. brevicollis*

The absence of teneurin genes in the complete and assembled genomes of the cnidarian *Nematostella vectensis* and placozoan *T. adhaerens* (that could be identified using the search methods employed to find teneurins in other metazoans) initially suggested that teneurins may have evolved about the time of the Cambrian radiation. However, during a routine search of predicted protein domain architectures that included RHS core-associated protein domains using the Pfam program, a sequence encoding EGF repeats, NHL domains, and YD repeats (in addition to the RHS core-associated protein domain) was identified in the genome of the choanoflagellate *Monosiga brevicollis*. Further analysis revealed that this predicted protein has the basic features of a metazoan teneurin: it is a type II transmembrane protein with a putative NLS in the ICD, eight EGF repeats, a cysteine-rich domain, and a C-terminal two-thirds with the same domain architecture as a prokaryotic YD-repeat protein (i.e., NHL domains, YD repeats, and an RHS core-associated protein domain; fig. 5*B*; supplementary table S3, Supplementary Material online). The predicted sequence (XP_001749414) is shown in its entirety in supplementary figure S1, Supplementary Material online, together with relevant alignments generated with ClustalW. The expression of the *M. brevicollis* teneurin is supported by two nonoverlapping ESTs (FE890769 and FE895158), both of which correspond to regions encoding the YD repeats.

The ICD of the *M. brevicollis* teneurin does not align significantly with the ICDs from other teneurins, and it lacks SH3-binding motifs. In addition, ProP fails to identify any potential furin cleavage sites in this teneurin. There are eight EGF repeats, but there are no free cysteines to support dimerization. Adjacent to the EGF repeats is a cysteine-rich region that is highly conserved: the exact 23aa consensus sequence ExxCx(D/N)xxDx(D/E)xDxxxDCxxx(D/E)CCxxxxCxxxxxC is found in all the teneurins analyzed except for *S. purpuratus* teneurin (which has one additional "x" between the fourth and fifth cysteine) and *S. mansoni* teneurin (which is missing the sixth cysteine). In fact, using the *M. brevicollis* cysteine-rich domain sequence in a tBLASTn search of all nucleotide sequences uncovers all the teneurins identified above that are listed on GenBank. However, neither this method nor the other search methods we employed to identify teneurin sequences revealed teneurins in sequences from sponges, placozoans, ctenophores, cnidarians, fungi, ichthyospores or nucleariids. Interestingly, a similar cysteine-rich sequence is found in an endo-1,3-beta-glucosidase from the diatom *Phaeodactylum tricornutum* (XP_002181321). This sequence is 46% identical and 71% similar to the 35aa cysteine-rich domain of *M. brevicollis* and it includes a core stretch of 23aa that is 65% identical and 91% similar (table 1). This 23aa core domain aligns well with the Cu^{++}-binding motif of MopE (Helland et al. 2008), and similar sequences are found in

Tucker et al. · doi:10.1093/molbev/msr271 MBE

Table 2. Alignment of Most Similar Sequences with YD Repeats from *Monosiga brevicollis* Teneurin[a]

M. brevicollis	YDVDGQLTQVLEDGAEVESYSYDVNGNRVAWNVRG---AAHSATYGADDA
Syntrophobacter fumaroxidans[b]	YDSLGRLLAVRLDGVPAEEYRYDVNGNRVEETNTPRGIIGRTSTYSEEDH
	** *:* * **. .*.* ******** . .:::**. :*
M. brevicollis	VFTVDGQSYAVDVDGFLTSVRG----MSLAYSGRGELLSATLPSGAGTVR
S. fumaroxidans	LLTSGGTVYRYDADGFLTTRTEGSAVTRYVYSSRGELLSVALPDGK-RIE
	::* .* * *.*****: .**.******.:**.* :.
M. brevicollis	YRYDGFGRRI
S. fumaroxidans	YVNDPLGRRI 46% identical/69% similar
	* * :****
M. brevicollis	YDVDGQLTQVLEDGAEVESYSYDVNGN-----RVAWNVRGAAHSATYGAD
Desulfococcus oleovorans[c]	YDEMGRLETVTKDGTLVESYSYDSTPYGTCTYQMNTLRGIAGRVLDYDAE
	** *:* * :**: ******* . :: *.: *.*:
M. brevicollis	DAVFTVDGQSYAVDVDGFLTSVRG----MSLAYSGRGELLSATLPSGAGT
D. oleovorans	DHLLSAGGTDYQYDLDGFLTSKTSGAETTYYDYSSRGELLSVDLPDGT-D
	* :::..* .* *:****** . **.******.* *.*:
M. brevicollis	VRYRYDGFGRRI
D. oleovorans	ITYVHDPLGRRI 43% identical/67% similar
	: * :* :****
M. brevicollis	YDVDGQLTQVLEDGAEVESYSYDVNGNRVAWNVRGAAHS--ATYGADDAV
Homo sapiens Ten-4	YDADGQLQTVSINDKPLWRYSYDLNGNLHLLSPGNSARLTPLRYDIRDRI
	.** * :. : ****:***. .:*: *. * :
M. brevicollis	FTVDGQSYAVDVDGFLTSVRGMSLAYSGRGELLSATLPSGAGTVRYRYDG
H. sapiens Ten-4	TRLGDVQYKMDEDGFLRQRGGDIFEYNSAGLLIKAYNRAGSWSVRYRYDG
	:.. .* :* **** . * : *.. * *:.* :*: :*******
M. brevicollis	FGRRI
	LGRRV 38% identical/63% similar
	:***:

[a] As determined by tBLASTn with aa2276–aa2379 of *M. brevicollis* teneurin against the NCBI Nucleotide Collection database and aligned with ClustalW 2.1.
[b] YD protein from genomic sequence CP000478 (4941818–4942148).
[c] YD protein from genomic sequence CP000859 (1374170–1373806).

the metal-binding region of a predicted gametolysin from *Volvox carteri* (table 1 [XP_002958497]) and *Chlamydomonas reinhardtii* (XP_001695639). Alignment of the 23aa motif reveals that the core of the cysteine-rich region of *M. brevicollis* teneurin is most similar to the diatom sequence and the core of the cysteine-rich region of trematode teneurin; the same regions from other metazoan teneurins are conserved but not to the same extent (table 1).

The NHL repeats and YD repeats of *M. brevicollis* align better with the YD-repeat proteins of some prokaryotes than with the YD repeats found in metazoan teneurins. The NHL domains are most similar (31% identical) to the YD-repeat protein of *Herpetosiphon aurantiacus* (ABX04679), a predatory filamentous chloroflex bacterium that lives in soil and freshwater. A stretch of 103aa corresponding to YD repeats 17-21 of *M. brevicollus* teneurin was analyzed further using tBLASTn and the entire NCBI nucleotide collection. This stretch is most similar to the YD-repeat proteins of *Syntrophobacter fumaroxidans* (a freshwater bacterium) and *Desulfococcus oleovorans* (which lives in coastal waters) and then to the YD repeats found in human teneurin-4 (table 2).

The *M. brevicollis* teneurin is predicted from sequences encoded on just four exons that are separated by introns of 129, 206 and 105 bp (in contrast, human teneurin-1 is encoded on 29 exons and the average intron is 8 kb). Remarkably, the region of the predicted protein corresponding to the four C-terminal EGF repeats, the cysteine-rich domain, the NHL repeats, the YD repeats, and the RHS core-associated protein domain is encoded on a single giant exon of 6829 bp (fig. 5*B*). For comparison, the

corresponding regions of human teneurin-1, *S. mansoni* teneurin, and *C. elegans* Ten-1L are encoded on 20, 21, and 7 exons, respectively (see also Minet and Chiquet-Ehrismann 2000).

Discussion

Here, we have used predictions based on proteomics to determine which teneurins may be processed such that the ECD becomes incorporated into the ECM and which teneurins may be processed such that the ICD is transported to the nucleus. Our predictions are validated by our previous experimental studies with avian teneurins. For example, we showed that a recombinant fusion protein with the ECD of chicken teneurin-2 was cleaved in vitro at a furin site between the transmembrane domain and the EGF repeats (Rubin et al. 1999). Consistent with this observation, we also showed that antibodies to the ECD of chicken teneurin-2 not only labeled the cell surface but also the ECM of chicken embryos (Tucker et al. 2001). When tagged chicken teneurin-2 ICD is overexpressed in HT1080 cells the recombinant ICD localizes to the nucleus (Bagutti et al. 2003), but there is no evidence published to date that the teneurin-2 ICD is processed and transported to the cell nucleus in vivo. In contrast, antibodies to the ECD of chicken teneurin-1 failed to stain the ECM, but antibodies to the ICD of teneurin-1 routinely stained the nuclei of cells in vitro and in histological sections of embryos (Kenzelmann et al. 2008). Moreover, when the sequence RKRK in the avian teneurin-1 ICD is mutated to AAAA it no longer localizes to the nucleus in vitro (Kenzelmann et al. 2008). Here, we show that the ICD of chicken teneurin-1 (specifically, the RKRK and flanking sequences) is predicted with a high likelihood to be located in the nucleus (NLS Mapper score = 9.0) and that the ICD of chicken teneurin-2 is much less likely to be nuclear (NLS Mapper score = 2.7). Similarly, the chicken teneurin-2 furin-cleavage site that we previously demonstrated to be functional is predicted by ProP to be active (score = 0.65), but no such site is found in chicken teneurin-1. Thus, teneurin-1 and/or teneurin-4 are most likely to be processed (by a yet unknown mechanism) so that the ICD can move to the nucleus, and teneurin-2 and/or teneurin-3 are more likely to have the ECD shed into the ECM. The shared features of these pairs of teneurins are also reflected in their predicted origins: teneurins-1 and -4 appear to have evolved from a gene duplication, as do teneurins-2 and -3 (fig. 3; see also Minet and Chiquet-Ehrismann 2000). All the chordate teneurins examined here except *D. rerio* teneurin-1 are likely to be cleaved near the C-terminus. This may be a step that precedes the formation of the teneurin-derived C-terminal neuropeptides characterized by others (reviewed by Rotzinger et al. 2010).

Using a yeast two-hybrid screen, Nunes et al. (2005) found that the SH3 domains of CAP/ponsin interact with the second proline-rich SH3-binding motif of chicken teneurin-1; the identical motif is present in teneurin-4. CAP/ponsin in turn binds to vinculin, which could anchor the ICD of teneurins to the actin cytoskeleton. A predicted SH3-binding motif at the same location in teneurin-2 does not bind CAP/ponsin even though it varies from the motif in teneurins-1 and -4 by only a single amino acid (Nunes et al. 2005). This led us to analyze teneurins from a broad range of taxa for SH3-binding motifs. The teneurin ICDs from each species examined, except for *S. mansoni* and *M. brevicollis*, contained one or more consensus SH3-binding motif. Interestingly, *S. mansoni* and *M. brevicollis* are the only species examined with teneurins lacking the capacity to dimerize. Perhaps dimerization is necessary for the ICD-interacting proteins to link teneurins to the cytoskeleton or to regulate the processes necessary for ICD nuclear localization.

Databases (e.g., GenBank, Ensembl, JGI, UniProt) contain listings for numerous teneurin variants. Most of these variants are based on predicted sequences, but some are based on cDNAs and ESTs. Here, we chose to study the range of alternative splicing in the ICD of human and chicken teneurin by PCR. The ICDs of human and chicken teneurins tend to be encoded on two pairs of neighboring exons separated by a large intron. Additional exons, which often are not conserved between birds and man and which frequently are subjected to alternative splicing, are sometimes found between the two pairs of exons. Alternatively spliced exons do not contain recognizable SH3-binding domains or NLSs, so the significance of these variations is not clear. Interestingly, an alternatively spliced exon in human teneurin-2 may represent an alternative start site, as ESTs with this sequence do not contain sequence from exons 1 or 2, and sequences encoded on this exon are not found in the PCR products amplified using a primer based on sequences found in exon 1. A similar method for generating teneurin splice variants was shown previously for *C. elegans* (Drabikowski et al. 2005).

The extraordinary diversity of teleost fish is commonly attributed to the duplication of their genome followed by the selective retention of certain duplicated genes (see Jozefowicz et al. 2003; Postlethwait et al. 2004; Volff 2005). This has been supported by studies of *Hox* genes (Kurosawa et al. 2006; Zou et al. 2007). In contrast, comparisons of *Sox* genes in the zebrafish *D. rerio* and the stickleback *G. aculeatus* (Cresko et al. 2003) show the mutual retention of *Sox9a* and *Sox9b*, albeit with subtle differences in their patterns of expression. Here, we show that the zebrafish has retained genes encoding teneurin-2A and teneurin-2B, whereas the stickleback has retained genes encoding teneurin-3A and teneurin-3B. Current models of selective gene retention in teleosts predict that genes are preserved following degeneration of regulatory elements and the partitioning of function between the duplicated gene products (Force et al. 1999). It is likely that a large, multifunctional protein like a teneurin would be selected in this way, and differential retention and expression could contribute to speciation.

Previously we speculated that the RHS proteins of bacteria, which share significant sequence homologies with the C-terminal portion of teneurins, may have evolved from horizontal gene transfer from a metazoan teneurin to a symbiotic or pathogenic prokaryote (Minet and Chiquet-

Ehrismann 2000). However, the presence of a teneurin gene in *M. brevicollis*, but not in the other nonmetazoan opisthokonts (e.g., fungi), suggests that the gene evolved in a choanoflagellate, and the horizontal gene transfer was from a prokaryote to a eukaryote instead of the other way around. Horizontal gene transfer between predatory *M. brevicollis* and their prokaryotic prey has been described previously. For example, Foerstner et al. (2008) reported that a nitrile hydratase is encoded in the *M. brevicollis* genome that is most closely related to enzymes from proteobacteria; the absence of this enzyme from other eukaryotic genomes strongly implies horizontal gene transfer from prokaryotic prey to eukaryotic predator. Over a hundred genes originating from haptophytes and diatoms have also been found in *M. brevicollis* (Nedelcu et al. 2008; Sun et al. 2010), indicating that gene transfer may be a relatively common occurrence in these organisms. In fact, this may explain the origin of the highly conserved cysteine-rich domain, which is nearly identical to part of an enzyme from the diatom *P. tricornutum* (and is also similar to an enzyme in *V. carteri*) but is only found in teneurins in metazoa. If this is the case, *M. brevicollis* teneurin originated as a fusion protein acquired by horizontal gene transfer from both a prokaryote and a diatom or algae.

Choanoflagellates are believed to be the closest living relatives of metazoansn (King and Carroll 2001; Philippe et al. 2004; King et al. 2008). The presence of teneurins (which have been shown to play roles in cell–cell and cell–ECM interactions in a variety of tissues) on the surface of an ancestral choanoflagellate may have facilitated the evolution of metazoan multicellularity and the development of complex tissues. Similar roles have been proposed for cadherins, which appear to have evolved in a choanoflagellate as well (Abedin and King 2008). The two cadherins of *M. brevicollis* are found in the microvilli that form the feeding collar that surrounds the base of the flagellum, which has led to the hypothesis that this family of proteins, which is indispensable in the formation of meaningful cell–cell contacts in animal tissues, evolved as a means of catching prey. Teneurins may have evolved to do something similar. YD-repeat proteins are found on the surface of aquatic bacteria, and in vitro studies with eukaryotic cells show that teneurin expression leads to increased cell–cell adhesion (Rubin et al. 2002). The acquisition of the carbohydrate-rich YD-repeat proteins from a prokaryote by a choanoflagellate may have improved "fishing" for bacterial prey in the feeding collar. It will be interesting to determine where *M. brevicollis* teneurin is expressed to test this hypothesis.

The lowest branches of the metazoan tree of life include the ancestors of sponges, ctenophores, and cnidarians. Therefore, it is puzzling that we were able to identify teneurins in a choanoflagellate and in all the available genomes of Bilateria (i.e., deuterostomes and protostomes) but not in modern sponges or cnidarians. It is possible that our search methods were insufficient to find them. More likely they are present in some of these organisms but not in the organisms with complete and well-assembled genomes like *N. vectensis*. It will be important to scrutinize newly sequenced and assembled sponge and cnidarian genomes for teneurin genes as they become available. Another possibility is that teneurins evolved in a relatively advanced common ancestor of protostomes and deuterostomes, after the evolution of sponges and cnidarians. In this scenario, the teneurin in *M. brevicollis* would have been acquired by horizontal gene transfer from metazoan-derived detritus and not a prokaryote. Evidence against this hypothesis includes the relative similarity of the core region of the cysteine-rich domain and a diatom enzyme and the YD repeats to YD-repeat proteins from bacteria, as well as the organization of the *M. brevicollis* teneurin gene: most of the ECD is encoded on a single huge exon, not unlike the YD-repeat proteins of prokaryotes, and unlike the ECD of metazoan teneurins. Others (King et al. 2008) have reported that the number of introns per gene in *M. brevicollis* is similar (6.6) to the number found in human genes (7.7), so the unusually large exon encoding the ECD of *M. brevicollis* teneurin argues for origins from a prokaryotic and not metazoan, horizontal gene transfer, and the subsequent loss of teneurins from the genomes of modern sponges and cnidarians.

Teneurins are phylogenetically conserved among Bilateria, where they have been demonstrated to play critical roles in pattern formation, the organization of the ECM, and the development of the nervous system. Genomic analysis reveals an ancient origin of teneurins in single-celled choanoflagellates that may have assembled teneurins via horizontal gene transfer from two of its prey: diatoms and prokaryotes. Thus, the talent for gene acquisition by an ancestral choanoflagellate, perhaps to diversify its metabolic pathways and improve its ability to capture prey, may have contributed to the development of multicellularity in metazoans.

Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank John Hess for his comments on the manuscript and for assistance with the cloning and sequencing of avian teneurin-3.

References

Abedin M, King N. 2008. The premetazoan ancestry of cadherins. *Science* 319:946–948.

Bagutti C, Forro G, Ferralli J, Rubin B, Chiquet-Ehrismann R. 2003. The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription. *J Cell Sci.* 116:2957–2966.

Baumgartner S, Chiquet-Ehrismann R. 1993. Tena, a *Drosophila* gene related to tenascin, shows selective transcript localization. *Mech Dev.* 40:165–176.

Baumgartner S, Martin D, Hagios C, Chiquet-Ehrismann R. 1994. Tenm, a *Drosophila* gene related to tenascin, is a new pair-rule gene. *EMBO J.* 13:3728–3740.

Beckmann J, Vitobello A, Ferralli J, Kenzelmann Broz D, Rijli FM, Chiquet-Ehrismann R. 2011. Human teneurin-1 is a direct target

of the homeobox transcription factor EMX2 at a novel alternate promoter. *BMC Dev Biol.* 11:35.

Cresko WA, Yan YL, Baltrus DA, Amores A, Singer A, Rodríguez-Marí A, Postlethwait JH. 2003. Genome duplication, subfunction partitioning, and lineage divergence: sox9 in stickleback and zebrafish. *Dev Dyn.* 228(3):480–489.

Drabikowski K, Trzebiatowska A, Chiquet-Ehrismann R. 2005. *ten-1*, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Dev Biol.* 282:27–38.

Foerstner KU, Doerks T, Muller J, Raes J, Bork P. 2008. A nitrile hydratase in the eukaryote *Monosiga brevicollis*. *PLoS One* 3:e3976.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.

Helland R, Fjellbirkeland A, Karlsen OA, Ve T, Lillehaug JR, Jensen HB. 2008. An oxidized tryptophan facilitates copper binding in *Methylococcus capsulatus*-secreted protein MopE. *J Biol Chem.* 283:13897–13904.

Jozefowicz C, McClintock J, Prince V. 2003. The fates of zebrafish *Hox* gene duplicates. *J Struct Funct Genomics.* 3:185–194.

Kay BK, Williamson MP, Sudol M. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14:231–241.

Kenzelmann D, Chiquet-Ehrismann R, Leachman NT, Tucker RP. 2008. Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo. *BMC Dev Biol.* 8:30.

King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci U S A.* 98:15032–15037.

King N, Westbrook MJ, Young SL, Kuo A, et al. (36 co-authors). 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451:783–788.

Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A.* 106:10171–10176.

Kowanetz K, Szymkiewicz I, Haglund K, Kowanetz M, Husnjak K, Taylor JD, Soubeyran P, Engstrom U, Ladbury JE, Dikic I. 2003. Identification of a novel proline-arginine motif involved in CIN85-dependent clustering of Cbl and down-regulation of epidermal growth factor receptors. *J Biol Chem.* 278:39735–39746.

Kurosawa G, Takamatsu N, Takahashi M, et al. (15 co-authors). 2006. Organization and structure of *hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* 370:75–82.

Leamey CA, Merlin S, Lattouf P, Sawatari A, Zhou X, Demel N, Glendining KA, Oohashi T, Sur M, Fässler R. 2007. Ten_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. *PLoS Biol.* 5:e241.

Levine A, Bashan-Ahrend A, Budai-Hadrian O, Gartenberg D, Menasherow S, Wides R. 1994. *Odd Oz*: a novel *Drosophila* pair rule gene. *Cell* 77:587–598.

Li H, Bishop KM, O'Leary DD. 2006. Potential target genes of EMX2 include Odz/Ten-M and other gene families with implications for cortical patterning. *Mol Cell Neurosci.* 33:136–149.

Lovejoy DA, Rotzinger S, Barsyte-Lovejoy D. 2009. Evolution of complementary peptide systems: teneurin C-terminal-associated peptides and corticotropin-releasing factor superfamilies. *Ann N Y Acad Sci.* 1163:215–220.

Mayer BJ. 2001. SH3 domains: complexity in moderation. *J Cell Sci.* 114:1253–1263.

Minet AD, Rubin BP, Tucker RP, Baumgartner S, Chiquet-Ehrismann R. 1999. Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene *ten-m*, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci.* 112:2019–2032.

Minet AD, Chiquet-Ehrismann R. 2000. Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli*. *Gene* 257:87–97.

Nedelcu AM, Miles IH, Fagir AM, Karol K. 2008. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *J Evol Biol.* 21:1852–1860.

Nunes SM, Ferralli J, Choi K, Brown-Luedi M, Minet AD, Chiquet-Ehrismann R. 2005. The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix. *Exp Cell Res.* 305:122–132.

Oohashi T, Zhou XH, Feng K, Richter B, Morgelin M, Perez MT, Su WD, Chiquet-Ehrismann R, Rauch U, Fassler R. 1999. Mouse ten-m/Odz is a new family of dimeric type II transmembrane proteins expressed in many tissues. *J Cell Biol.* 145:563–577.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol.* 21:1740–1752.

Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 20:481–490.

Rakovitsky N, Buganim Y, Swissa T, Kinel-Tahan Y, Brenner S, Cohen MA, Levine A, Wides R. 2007. *Drosophila Ten-a* is a maternal pair-rule and patterning gene. *Mech Dev.* 124:911–924.

Rotzinger S, Lovejoy DA, Tan LA. 2010. Behavioral effects of neuropeptides in rodent models of depression and anxiety. *Peptides* 31:736–756.

Rubin BP, Tucker RP, Martin D, Chiquet-Ehrismann R. 1999. Teneurins: a novel family of neuronal cell surface proteins in vertebrates, homologous to the *Drosophila* pair-rule gene product Ten-m. *Dev Biol.* 216:195–209.

Rubin BP, Tucker RP, Brown-Luedi M, Martin D, Chiquet-Ehrismann R. 2002. Teneurin 2 is expressed by the neurons of the thalamofugal visual system in situ and promotes homophilic cell-cell adhesion in vitro. *Development* 129:4697–4705.

Sun G, Yang Z, Ishwar A, Huang J. 2010. Algal genes in the closest relatives of animals. *Mol Biol Evol.* 27:2879–2889.

Trubiani G, Al Chawaf A, Belsham DD, Barsyte-Lovejoy D, Lovejoy DA. 2007. Teneurin carboxy (C)-terminal associated peptide-1 inhibits alkalosis-associated necrotic neuronal death by stimulating superoxide dismutase and catalase activity in immortalized mouse hypothalamic cells. *Brain Res.* 1176:27–36.

Trzebiatowska A, Topf U, Sauder U, Drabikowski K, Chiquet-Ehrismann R. 2008. *Caenorhabditis elegans* teneurin, ten-1, is required for gonadal and pharyngeal basement membrane integrity and acts redundantly with integrin ina-1 and dystroglycan dgn-1. *Mol Biol Cell.* 19:3898–3908.

Tucker RP, Chiquet-Ehrismann R. 2006. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev Biol.* 290:237–245.

Tucker RP, Chiquet-Ehrismann R, Chevron MP, Martin D, Hall RJ, Rubin BP. 2001. Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced in vitro and in situ by FGF8. *Dev Dyn.* 220:27–39.

Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* 94:280–294.

Youderian P, Hartzell PL. 2007. Triple mutants uncover three new genes required for social motility in *Myxococcus xanthus*. *Genetics* 177:557–566.

Young TR, Leamey CA. 2009. Teneurins: important regulators of neural circuitry. *Int J Biochem Cell Biol.* 41:990–993.

Zou SM, Jiang XY, He ZZ, Yuan J, Yuan XN, Li SF. 2007. Hox gene clusters in blunt snout bream, *Megalobrama amblycephala* and comparison with those of zebrafish, fugu and medaka genomes. *Gene* 400:60–70.

4.2. Human Teneurin-1 is a direct target of the homeobox transcription factor EMX2 at a novel alternate promoter

Jan Beckmann, Antonio Vitobello, Jacqueline Ferralli, Daniela Kenzelmann Broz, Filippo M. Rijli and Ruth Chiquet-Ehrismann

BMC Developmental Biology 2011, 11:35

First published online: June, 8th 2011

My contribution to this paper:

For this study, I performed the 5'-RACE of human teneurin-1 and the subsequent sequence analysis leading to the discovery of the novel alternate promoter. I cloned all constructs used in the study and established all stable cell lines. Together with Jacqueline Ferralli, I planned and analyzed the promoter assays. Furthermore, I performed the gel shift assays to show direct binding of Emx2 to the homeobox binding site in the novel promoter and performed the QPCR analysis of teneurin-1 expression. Finally, I wrote the first draft of the manuscript and the final version with the input of the co-authors.

Beckmann et al. BMC Developmental Biology 2011, 11:35
http://www.biomedcentral.com/1471-213X/11/35

BMC Developmental Biology

RESEARCH ARTICLE Open Access

Human teneurin-1 is a direct target of the homeobox transcription factor EMX2 at a novel alternate promoter

Jan Beckmann[1,2†], Antonio Vitobello[1,2†], Jacqueline Ferralli[1], Daniela Kenzelmann Brož[3], Filippo M Rijli[1,2] and Ruth Chiquet-Ehrismann[1,2*]

Abstract

Background: Teneurin-1 is a member of a family of type II transmembrane proteins conserved from *C.elegans* to vertebrates. Teneurin expression in vertebrates is best studied in mouse and chicken, where the four members teneurin-1 to -4 are predominantly expressed in the developing nervous system in area specific patterns. Based on their distinct, complementary expression a possible function in the establishment of proper connectivity in the brain was postulated. However, the transcription factors contributing to these distinctive expression patterns are largely unknown. Emx2 is a homeobox transcription factor, known to be important for area specification in the developing cortex. A study of Emx2 knock-out mice suggested a role of Emx2 in regulating patterned teneurin expression.

Results: 5'RACE of human teneurin-1 revealed new alternative untranslated exons that are conserved in mouse and chicken. Closer analysis of the conserved region around the newly identified transcription start revealed promoter activity that was induced by EMX2. Mutation of a predicted homeobox binding site decreased the promoter activity in different reporter assays *in vitro* and *in vivo* in electroporated chick embryos. We show direct *in vivo* binding of EMX2 to the newly identified promoter element and finally confirm that the endogenous alternate transcript is specifically upregulated by EMX2.

Conclusions: We found that human teneurin-1 is directly regulated by EMX2 at a newly identified and conserved promoter region upstream of the published transcription start site, establishing teneurin-1 as the first human EMX2 target gene. We identify and characterize the EMX2 dependent promoter element of human teneurin-1.

Background

Many transmembrane proteins mediate cell-cell interactions and thereby regulate key developmental processes. Teneurins are a unique family of type II transmembrane proteins conserved from *Drosophila melanogaster* and *Caenorhabditis elegans* to vertebrates, where four paralogues exist called teneurin 1-4 [1]. This protein class was discovered in a screen for the *Drosophila* homologue of the extracellular matrix protein tenascin-C [2]. Structure and domain architecture are highly conserved across phyla. All proteins of the teneurin family share a large extracellular domain with eight tenascin-type EGF-like repeats followed by a region of conserved cysteines and YD repeats [3]. Recently, several publications suggested that the C-terminal parts of the teneurin proteins contain peptides with similarities to corticotrophin-releasing factor (CRF) and might have a function in modulating CRF-mediated behavior [4]. All vertebrate teneurins have an N-terminal intracellular domain with two polyproline motifs, EF-hand-like metal ion binding sites and several putative phosphorylation sites. This intracellular domain was shown to be cleaved from the membrane and translocates into the nucleus where it can interact with transcription factors and alter gene expression [5-7].

In *C. elegans*, RNAi knockdown and deletion of its single teneurin gene (Ten-1) results in a broad range of phenotypes, including defects in axon guidance and neuronal

* Correspondence: ruth.chiquet@fmi.ch
† Contributed equally
[1]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, CH-4058 Basel, Switzerland
Full list of author information is available at the end of the article

pathfinding, as well as gonadal disintegration and protrusion of the vulva [8-10]. *Drosophila* harbors two teneurin genes, Ten-a [2] and Ten-m/Odz [11,12]. Mutations in either of these genes result in embryonic lethality and Ten-a mutants enhance the segmentation phenotype of weak alleles of Ten-m/Odz [13]. It was also shown that teneurin expression is required for the proliferation and cellular identity in the *Drosophila* eye [14]. Extensive localization studies in mouse [15-17] and chicken [5,18-20] embryos, as well as in rat [21] and zebrafish [22] revealed that the different members of the teneurin protein family are expressed with overlapping patterns by distinct subpopulations of neurons. Experiments *in vitro* and *in vivo* showed that the different members of the teneurin family form disulfide-linked dimers [16,23] and promote homophilic cell-cell adhesions and neurite outgrowth [18,24]. These functions of the protein are believed to mediate correct pathfinding and area recognition of neurons. This was shown in the teneurin-3 knockdown mouse, which exhibits dramatic changes in the mapping of ipsilateral retinal inputs causing mismatches in binocular mapping. This is associated with major deficits in the performance of visually mediated behavioral tasks [25].

Recent findings suggest an important role for the teneurin protein family in establishing cortical arealization and patterning in the developing embryo. Teneurin-2 was found to be expressed in developing limbs, somites and craniofacial mesenchyme in a pattern strikingly similar to that of fibroblast growth factor 8 (Fgf8) and Fgf8 coated beads implanted into chicken limb buds induced ectopic teneurin-2 expression *in situ* [20]. Furthermore, teneurin-4 transcripts are down regulated, and the expression patterns of teneurins are shifted in the cortices of mice deficient in Emx2 [26]. These findings link the regulation of teneurin expression to Fgf8 and Emx2, two proteins that are part of a complex network of growth and transcription factors regulating arealization of the developing brain, a crucial event regulating sensory perception, the control of our movements and behavior (reviewed in [27]). The best studied protein in this network is Emx2. Emx2 is the vertebrate homologue of the *Drosophila* empty spiracles (*ems*) protein, which is involved in the development of the fly head [28]. This protein is a homeobox-containing transcription factor implicated in mouse cerebral cortex development [29]. It is expressed in a graded manner from rostral (low) to caudal (high) [30-33]. Knock-out and overexpression studies of Emx2 showed the function of this transcription factor in establishing the correct size and positioning of cortical areas [reviewed in 34]. Comparing expression analyses of different embryonic stages to the adult for both Emx2 [32,33] and teneurins [5,7,35] showed that areas of Emx2 expression (e.g., the cortical plate, dentate gyrus and the olfactory bulb) strongly correlate with areas of teneurin expression, suggesting a possible role of teneurins in mediating arealization.

The human teneurin-1 gene resides on the × chromosome at position Xq25, a locus with low gene density [reviewed in 36]. Beside severe mental retardation, patients suffering from a syndrome mapped to this locus also suffer from motor sensory neuropathy, deafness and severely impaired vision [37-41]. Given the predominant expression in the developing brain and its function in establishing proper connectivity in the brain, teneurin-1 is a potential target gene for causing XLMR.

In order to provide the basis for an investigation of possible deletions and mutations in teneurin-1 of XLMR patients, we decided to delineate the gene locus and determined the transcription start site(s) of human teneurin-1. We identified a novel promoter upstream of the published transcription start, which is conserved in chicken and mice. We show that EMX2 directly binds to and regulates human teneurin-1 expression at this alternate promoter.

Results

Identification of alternate transcription start sites of the teneurin-1 gene

Whereas the expression and localization of the different members of the teneurin protein family are well characterized, promoter regions regulating teneurin gene expression in vertebrates have not yet been studied. To find the transcription start point of human, mouse and chicken teneurin-1 (gene name is ODZ1), we performed 5'-RACE on brain cDNAs of the respective species. We used gene-specific primers derived from the first coding exon and in each case identified two classes of products (Figure 1 and Table 1). The first class ended with the 5'UTR of the published first exon containing the translation start site (as depicted in the genome browser as ODZ1), and the second class included additional non-coding exons. Two additional exons were found in human, three in mouse, and one in chicken teneurin-1. All of these exons were between 80 kb and over 200 kb distant from the first coding exon. CpG islands were found surrounding the newly identified alternate first exon suggesting promoter activity in this region. Using 4 kb of sequence surrounding the newly identified first exon of human teneurin-1 to BLAST the mouse genome revealed that this entire region was conserved between species with an overall sequence identity of 58%, and included local sequence identities of over 90%. Based on these findings, we considered that teneurin-1 expression is regulated by two different promoters that are used to differentially regulate teneurin-1 expression.

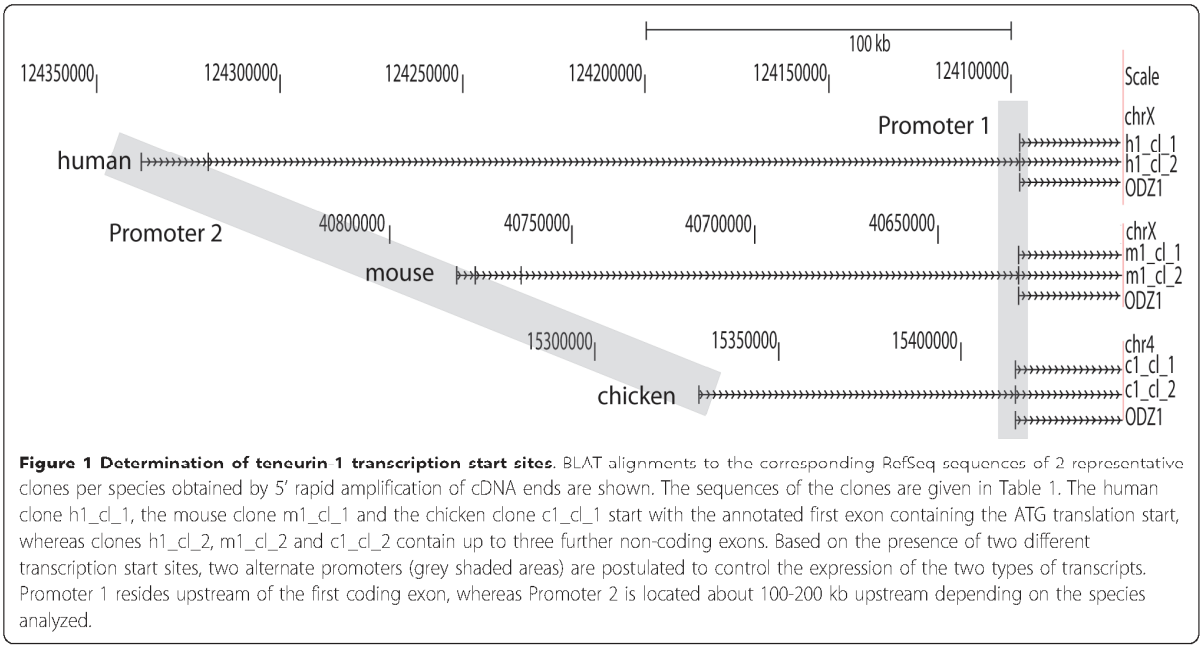


Figure 1 Determination of teneurin-1 transcription start sites. BLAT alignments to the corresponding RefSeq sequences of 2 representative clones per species obtained by 5' rapid amplification of cDNA ends are shown. The sequences of the clones are given in Table 1. The human clone h1_cl_1, the mouse clone m1_cl_1 and the chicken clone c1_cl_1 start with the annotated first exon containing the ATG translation start, whereas clones h1_cl_2, m1_cl_2 and c1_cl_2 contain up to three further non-coding exons. Based on the presence of two different transcription start sites, two alternate promoters (grey shaded areas) are postulated to control the expression of the two types of transcripts. Promoter 1 resides upstream of the first coding exon, whereas Promoter 2 is located about 100-200 kb upstream depending on the species analyzed.

EMX2 transactivates teneurin-1 promoter reporter constructs in cell culture

To test whether human teneurin-1 is a direct target gene of EMX2, we set up a reporter gene assay. We obtained a myc-flag-tagged EMX2 expression plasmid to transfect NIH3T3 cells. Recombinant EMX2 could be detected in cell extracts as a 37kD protein band on Western blots with a FLAG antibody (Figure 2a) and the protein accumulated in the nuclei of the cells as shown by immunostaining (Figure 2b). In order to do promoter reporter assays, we cloned a 4 kb fragment of highly conserved genomic sequence around the published transcription start site, as well as around the newly determined upstream transcription start site of human teneurin-1 into a pSEAP2-basic reporter vector. These promoter reporter constructs were co-transfected with the EMX2 plasmid into HEK293 cells and reporter gene activity was measured (Figure 2c). Interestingly, EMX2 was able to strongly induce reporter gene activity from the newly identified upstream promoter 2, but not from promoter 1, which remained unchanged compared to the empty vector control. Previously it was shown that EMX2 binds to a homeobox binding motif in the Wnt-1 promoter [42]. Upon sequence analysis of the promoter 2 construct we found one conserved site with a high score for EMX2 binding, while the promoter 1 construct possessed several high scoring binding sites. This indicates that the mere presence of core sequences of homeobox binding elements is not sufficient *per se* for the induction by EMX2, but the context may matter as well. To examine whether the homeobox binding motif in the promoter 2 construct contributes to the reporter gene activation upon EMX2 co-transfection, we mutated this motif and measured secreted embryonic alkaline phosphatase (SEAP) reporter gene activity. Indeed, the SEAP activity dropped significantly to 50% compared to the wild-type construct (Figure 2c).

EMX2 transactivates a teneurin-1 promoter construct in chick embryos electroporated *in ovo*

To further prove the promoter activity of the upstream sequence, as well as its dependence on EMX2 expression, we carried out an *in ovo* reporter gene assay in chick embryos. Upon co-electroporation of the promoter 2-lacZ construct with the EMX2 expression vector in the chick embryo neural tube, lacZ staining was strongly detected in the electroporated area (Figure 3b), whereas no staining was visible with the empty lacZ vector control (data not shown) or upon electroporation with the promoter 2-lacZ construct alone (Figure 3b). A GFP-expressing construct was always co-electroporated as well and used as a control for the efficiency of electroporation (Figure 3a).

To confirm the influence of the putative EMX2 binding motif on reporter gene activation *in vivo*, we co-electroporated the mutant promoter 2-lacZ construct with EMX2 and compared the staining with that of embryos co-electroporated with the wild-type promoter 2-lacZ construct and EMX2. Similar to the cell culture based reporter gene assay (Figure 2), reporter gene activity was also strongly reduced *in vivo*. Indeed, the vast majority of the embryos co-electroporated with EMX2 and the mutated promoter 2-lacZ construct showed a much fainter lacZ staining than those co-electroporated with

Table 1 Sequences obtained in 5'RACE (Translation start site in bold and underlined)

clone name	sequence
h1_cl_1	TTTTTTTTTTTTTTTGAACTGAGCTTGCTTAATCAGAG**ATG**GAGCAAACTGACTGCAAACCCTACCA GCCTCTACCAAAAGTCAAGCATGAAATGGATCTAGCTTACACCAGTTCTTCTGATGAGAGTGAAG ATGGAAGAAAACCAAGACAGTCATACAACTCCAGGGAGACCCTGCACGAGTATAACCAGGAGCT GAGGATGAATTACAATAGCCAGAGTAGAAAGAGGAAAGAAGTAGAAAAATCTACTCAAGAGAT GGAATTCTGTGAAACCTCTCACACTCTGTGCTCTGGCTACCAAACAGACATGCACAGCGTTTCTCG GCATGGCTACCATCTAGA
h1_cl_2	TTTTTTTTTTTTTTTGGCGGGGGANCAGCACCTGGGGACGCCGCCGAAACTTGCGCTTGGAATA GGAATTACAAGGGTGACCTTTATTCCGCTGTCTCCTTTTTGATTCCCATAACTTCTGGACCTATCAA GGACTGCTTGCATTAAAGGACTTCCTCATCCTTTTTTTTCATGAAACTGAGCTTGCTTAATCAGAG**A** **TG**GAGCAAACTGACTGCAAACCCTACCAGCCTCTACCAAAAGTCAAGCATGAAATGGATCTAGCT TACACCAGTTCTTCTGATGAGAGTGAAGATGGAAGAAAACCAAGACAGTCATACAACTCCAGGG AGACCCTGCACGAGTATAACCAGGAGCTGAGGATGAATTACAATAGCCAGAGTAGAAAGAGGA AAGAAGTAGAAAAATCTACTCAAGAGATGGAATTCTGTGAAACCTCTCACACTCTGTGCTCTGGC TACCAAACAGACATGCACAGCGTTTCTCGGCATGGCTACCATCTAGA
m1_cl_1	TTTTTTCCACCGCCACCTCCTCCACATGCCTGCACCTGTGCCAGGAAGCCACCTCCTACAGTGGAC TCTCTACAAAGAAGATCAATGACTACCCGCAGCCAGCCCAGCCCAGCTGCTCCTGCTCCTCCAACC AGCACACAGGATTCGGTTCATCTGCATAACAGCTGGGTCTTGAACAGTAACATACCGCTGGAGAC CAGGTACATTTTATGATTGACCATTTCAGCAAAGACTGTTTTCATTAAAGAACTTCCTTATCCTTTT TTCATGAAACTCAGCTTGCTTAATCAGAG**ATG**GAGCAAACAGACTGCAAACCTTATCAGCCTCTG TCCAAAGTCAAGCATGTCTAGA
m1_cl_2	TTTTTTCCCGCAGGAACCAGCAAAGACGCCCTAAGTCCAGCGCACTTACAGCACACCAGCAGAGC TGAGTACCTGGCAAGGAGGCGGGGGACCGCACCTGAGGACATCACTGAAACTTGCGCCTGGACT AGTCCTTCTACTGCCATGGAAACTAGATGGCACAGACAGCGGAGAGTCACTCATTCAGAACAGG GGCCCCCTTTTTAATTTCATGTCAGCCTGTTGGTCCCTGAAAGTAACTGAAAAGGAATTACAAGA GCGACTTTTATTCTGTGTAACTTCTCTTCTGGATCTAACAAGGTACATTTTATGATTGACCATTTCA GCAAAGACTGTTTTCATTAAAGAACTTCCTTATCCTTTTTTCATGAAACTCAGCTTGCTTAATCAGA G**ATG**GAGCAAACAGACTGCAAACCTTATCAGCCTCTGTCCAAAGTCAAGCATGTCTAGA
c1_cl_1	TTTTTTCCTCATTCCTTAAGGAATTCCAGTTGCTTGTTTTCATGATTTTGAGCCTATTCAGCCAGA G**ATG**GAGCAGATGGACTGCAAACCCTACCAGCCACTGTCAAAAGTTAAACATGAAGTGGATCTA ACNTTACACAAGTTCTTCAGATGAAAGTGAAGATGGCAGAAAGCAAAGGCAATCTTATGACTCA AGAGAAACTCTGAATGAATATAGCCAAGAGCTAAGACTGAACTACAACAGTCAAGGCAGAAAAA GAAAAAATACTGACCAATCCACACAAGACATGGAATTCTGTGAGACACCCCACATTCTGTGCTCT GGCTACCAAACAGATTTACATGGTGTGTCGGAGCACAGCTACCCACTAGAGGTGGGCTCAGATG TTGATACTGAAACCGAAGGTGGCGCATCACCAGATCATGCCCTGAGGATGTGGATGAGGGGGAT GAAGTCAGAACACAGCTCCTGTCCGTCAAGCCGGGCAAACTCAGCGTTGTCCCTGACTGACACTG ACCATGAGAGGAAGTCTGATGGGGAGAATGACATGCCGGGGAGCCCACACAACCAGTTCACGTT TCTAGA
c1_cl_2	TTTTTTTCGCCGAGCCTAGAGGCGATGGGAGCTGCCGAGCCGGGGCGCTGCTGAAAGTTCAGCC GGTGGCCGCGCAGCGCGGACTCATCCTTAAGGAATTCCAGTTGCTTGTTTTCATGATTTTGAGCC CATTCAGCCAGAG**ATG**GAGCAGATGGACTGCAAACCCTACCAGCCACTGTCAAAAGTTAAACAT GAAGTGGATCTAACTTACACAAGTTCTTCAGATGAAAGTGAAGATGGCAGAAAGCAAAGGCAAT CTTATGACTCAAGAGAAACTCTGAATGAATATAGCCAAGAGCTAAGACTGAACTACAACAGTCAA AGCAGAAAAAGAAAAAATACTGACCAATCCACACAAGACATGGAATTCTGTGAGACACCCCACA TTCTGTGCTCTGGCTACCAAACAGATTTACATGGTGTGTCGGAGCACAGATACTCTAGA

EMX2 and the wild type promoter 2-lacZ construct (Figure 3b). These findings show that EMX2 is able to induce reporter gene activity at the alternate teneurin-1 promoter and that this activation is greatly dependent on an intact homeobox binding motif.

EMX2 binds a homeobox core element in the alternate teneurin-1 promoter

To prove that the activation of the construct is due to the direct binding of EMX2 to the homeobox binding motif in the novel upstream promoter, we performed an electrophoretic mobility shift assay (EMSA). Nuclear extracts of HEK293 cells transfected with the EMX2 construct showed a shift of the labeled probe containing the putative homeobox binding site of the upstream promoter, whereas no shift was observed in nuclear extracts of untransfected HEK293 cells or with a mutated labeled probe (Figure 4a, lanes 1-3). To show the specificity of the binding, the effect of wild-type or mutated unlabeled oligo-nucleotide on the protein/DNA interaction was analyzed. Whereas no shift of the oligo-probe was visible when competing with the unlabeled wildtype oligo-nucleotide, the shift was still detected in the presence of excess mutant unlabeled oligo-nucleotide (compare lanes 4 and 5). The complex was super-shifted with a c-myc antibody against the tagged EMX2, and indeed the shifted band disappeared (lane 6). This indicates a direct binding of EMX2 to the probe, but due to a high unspecific background the super-shifted band could not be resolved.

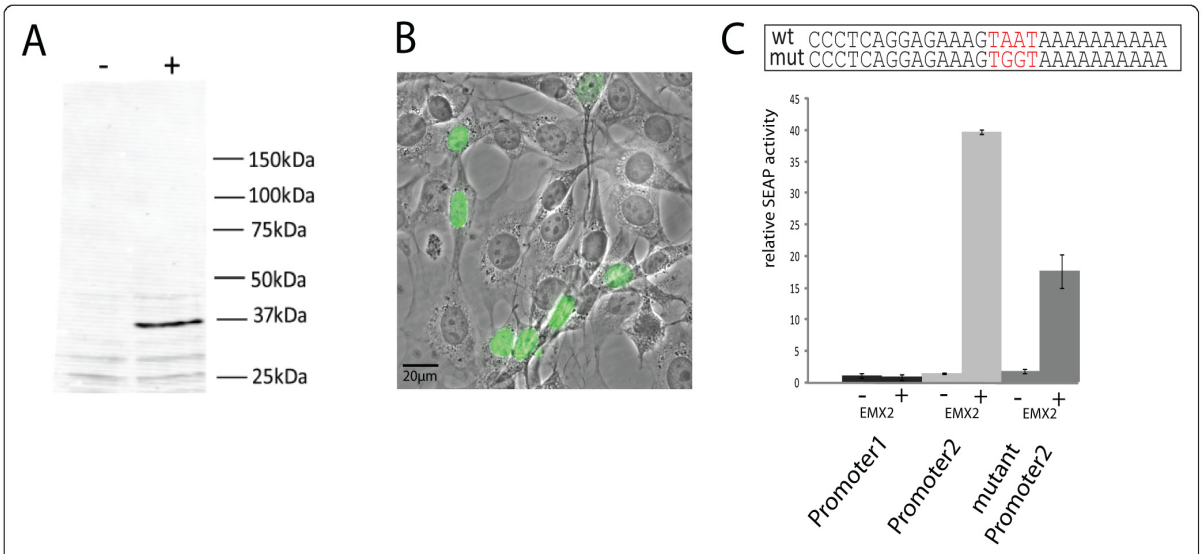


Figure 2 Transcriptional activation of teneurin-1 promoter reporter constructs by EMX2. (A) NIH3T3 cells were transfected with a flag-myc-tagged EMX2 construct or an empty vector and tested for expression by Western blot of whole cell extract with anti-Flag. (B) Immunostaining of the cells with anti-Flag reveals the nuclear accumulation of the EMX2 protein in the nuclei of the cells. (C) HEK293 cells were transfected with EMX2 or empty vector control and co-transfected with different promotor constructs as indicated. The graph shows relative SEAP values for one representative experiment performed in triplicates (n = 5). The activity obtained with the Promotor 1 construct is arbitrarily set to 1. Error bars display standard deviation of the mean. The sequences of the region harboring the putative binding site and the mutation used (mut Promoter 2) are shown in the box above the bar graph.

To reduce the unspecific background obtained with the nuclear extract, we tested EMX2 produced by *in vitro* transcription and translation in the gel shift assay. Whereas no shift of the labeled probe was detected with mock extracts, the same shifted band as with the nuclear extract could be detected with *in vitro* transcribed and translated EMX2 (Figure 4b, lanes 1 and 2), while unspecific background was greatly reduced. The binding of the protein to the probe was successfully competed with an excess of unlabeled wildtype oligo-nucleotides (Figure 4b, lane 3), whereas no competition was detected for unlabeled mutated oligo-nucleotides (Figure 4b, lane 4). Adding c-myc antibody to the binding reaction resulted in a super-shifted band (Figure 4b, lane 5), indicating a direct binding of EMX2 to the homeobox motif in the alternate teneurin-1 promoter.

To test whether an interaction between EMX2 and the binding site in the teneurin-1 promoter can also occur *in vivo* without overexpression of the EMX2 protein, we tested nuclear extracts of brains from E12.5 embryos known to express high EMX2 levels in the EMSA assay (Figure 4c). We were able to detect a shift of the band with the embryo extract, which runs lower than the complex of the overexpressed tagged protein in the control (compare Figure 4c, lanes 1, 2). We were able to compete the binding to the probe with wildtype unlabeled oligo-nucleotides, whereas no competition was detected using the unlabeled mutated oligo-nucleotide (Figure 4c, lanes 3-4). As a final proof of direct binding of EMX2 to the endogenous teneurin-1 promoter at the homeobox binding site *in vivo*, we performed chromatin immunoprecipitation (ChIP) in chicken embryos electroporated with the FLAG-myc-tagged EMX2 construct. Electroporations were performed in developing telencephalic regions in order to test the ability of Flag-myc-tagged EMX2 to bind the target region in its physiological cell context. Indeed, we detected specific enrichment of the target region containing the homeobox binding site after ChIP with the anti-FLAG antibody, recognizing the electroporated tagged EMX2 protein compared to a negative control region, which was not the case in control embryos (Figure 4d).

Teneurin-1 expression pattern correlates with that of *EMX2* in E14.5 embryos

To test whether the endogenous teneurin-1 expression pattern overlaps with sites of EMX2 expression in the developing brain, we performed *in situ* hybridizations with a probe for EMX2, a probe for total teneurin-1 and an additional probe specific for the alternate transcript of teneurin-1 on adjacent sagittal brain sections (Figure 5). The staining for teneurin-1 transcripts showed expression at sites that are in accordance with those reported before for E15.5 embryos [26]. Interestingly, staining with a probe specific for the alternate transcript revealed the same staining pattern as the probe for total teneurin-1, indicating that the long transcripts are indeed expressed at these stages of embryogenesis. We detected a correlation

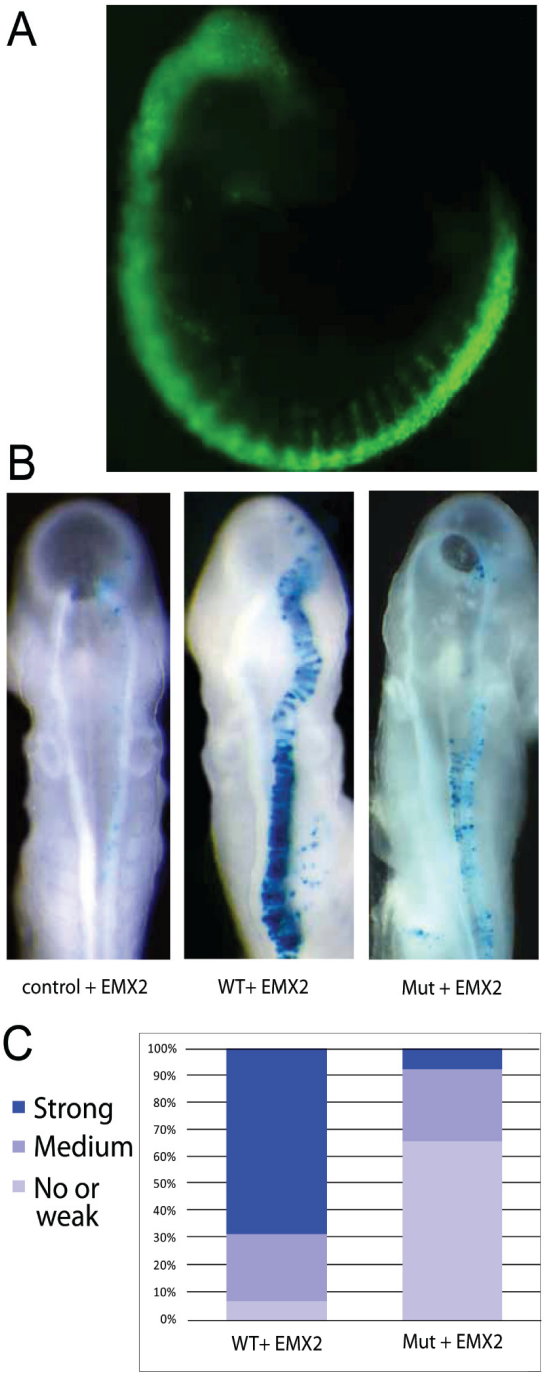


Figure 3 Chicken Embryo Electroporation. (A) Embryos were electroporated at Hamburger and Hamilton stage 10 and collected 24 h later. A CMV-GFP plasmid was always used as positive control of the electroporation and only embryos showing an expression pattern as shown were kept. Embryos electroporated with beta-globin_LacZ (BGZ40) plasmid alone or in combination with EMX2 in pCMV6-Entry (CMV-EMX2) were used as negative controls (not shown). (B) A 4 kb genomic fragment of the teneurin-1 promoter was cloned into the reporter plasmid beta-globin-LacZ (ten-1BGZ40) and electroporated alone (WT control) or in combination with CMV-EMX2 plasmid (WT + EMX2). A mutated version of ten-1BGZ40 lacking a potential EMX2 binding site (mut ten-1BGZ40) was electroporated alone (data not shown) or in combination with CMV-EMX2 (Mut + EMX2). (C) Electroporation results are summarized in stacked columns. Data are represented as percentile of the total number of electroporated embryos (n = 16 for the WT construct and n = 26 for the mutated construct) and are classified according to three levels of reporter expression: strong staining, medium staining and no or weak staining.

between the teneurin-1 signals and the EMX2 signal in a caudal high to rostral low gradient. We find teneurin-1 being expressed in the marginal, but not in the ventricular zone of the cortex. Especially good correlations were found in the caudal cortex, olfactory bulb (ob) and hippocampus (hi) (Figure 5).

EMX2 specifically induces the transcription of the alternate transcript

To test whether EMX2 is able to induce the endogenous teneurin-1 gene from the alternate promoter, we set up a real-time Q-PCR assay. We compared the mRNA expression level for total teneurin-1, as well as for the presence of the exons specific for the alternate transcript in parental HEK293-ECR cells with HEK293-ECR cells stably expressing myc-flag-tagged EMX2 (Figure 6). Indeed, the EMX2 expressing cells showed significantly ($p < 0.01$) elevated transcript levels of total teneurin-1. Although we generally observed a low expression level for the alternate transcript in our cells, it showed a much higher fold induction upon EMX2 expression than the total mRNA. This is further support for our reporter gene studies on the level of the endogenous gene and represents an independent confirmation that EMX2 specifically acts on the alternate promoter of teneurin-1.

Discussion

In this work, we characterized the teneurin-1 gene locus and found novel upstream exons which are conserved between species. These new exons expand the size of the Odz1 locus to more than 800 kb, harboring one intron which is more than 200 kb in size. Genes with large introns have been reported before [43]. A continuous transcription of the entire gene, given a polymerization rate of 3800 nucleotides per minute by RNA polymerase II, would take 3.5 h [44]. This might add another level of regulation of the defined expression in time and space. Here we show that there are at least two promoters regulating teneurin-1 expression with one alternate promoter upstream of the published transcription start. Only this alternate promoter was inducible by EMX2 in reporter gene assays and cells stably overexpressing EMX2 exhibited an increase of the resulting alternate transcript. A single homeobox binding site seems to be critical for the

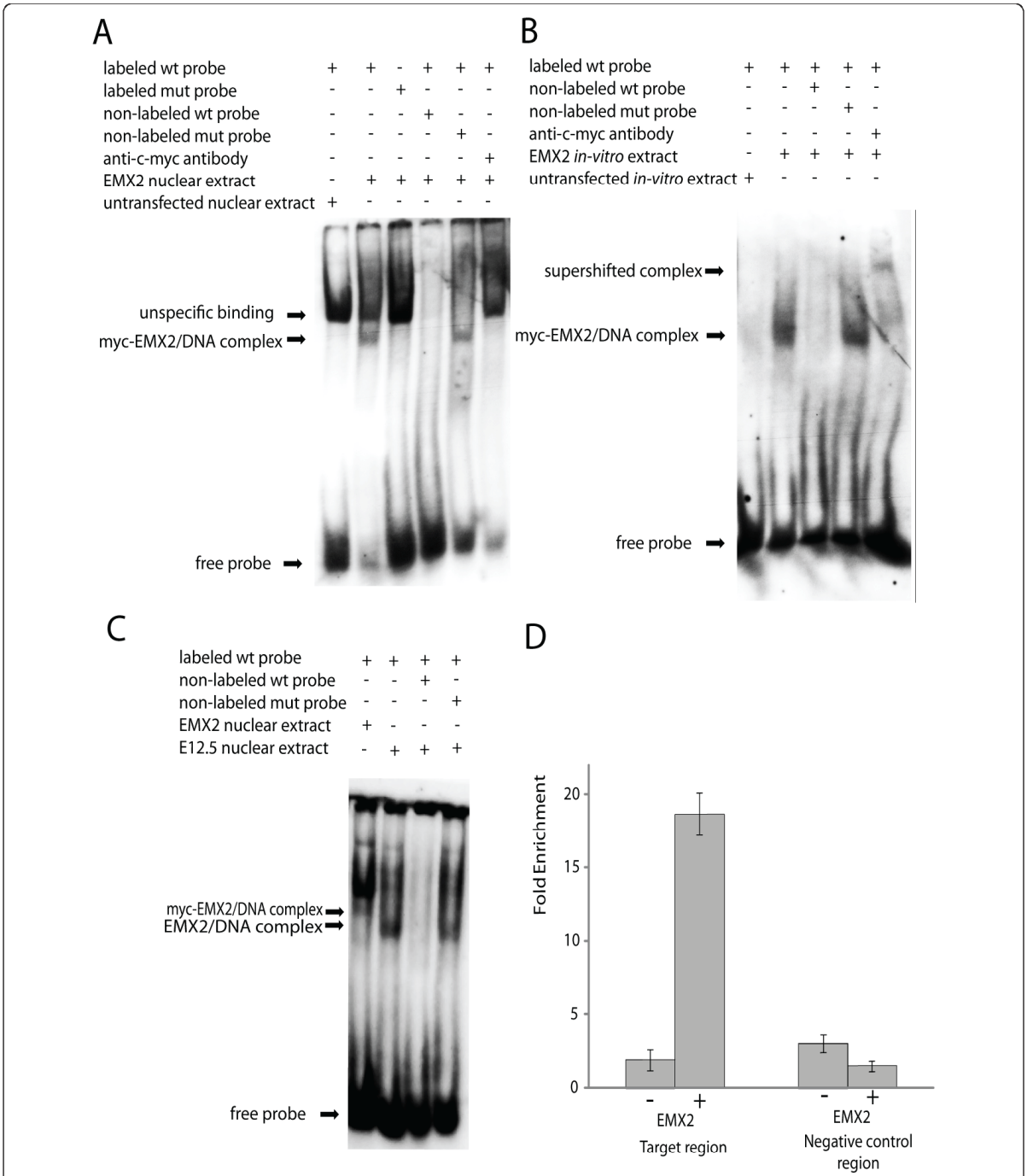


Figure 4 Direct binding of EMX2 to teneurin-1 promoter oligo-nucleotide probes. (A) Electrophoretic Mobility Shift Assays (EMSA) using EMX2 containing nuclear extract or control nuclear extract were performed in the presence or absence of Dig-labeled probe, unlabeled probe for competition, mutated probes and anti-myc antibody as indicated above the lanes. A specific myc-EMX2/DNA complex could be detected as indicated by an arrow. (B) EMSA with in vitro transcribed and translated EMX2 protein or control extracts were analyzed using Dig-labeled wildtype and mutated probes as indicated. Binding to the probe resulted in a myc-EMX2/DNA complex as indicated by an arrow that was competed by unlabeled probe and resulted in a supershifted complex after addition of anti-myc-antibody as indicated. (C) EMSAs with nuclear extracts of E12.5 embryos and nuclear extract of myc-EMX2 overexpressing cells as a control were performed using Dig-labeled wildtype probe. Binding to the probe resulting in a myc-EMX2/DNA complex and an EMX2/DNA complex is indicated by arrows. (D) ChIP of chicken embryos electroporated with flag-myc-tagged EMX2 (+) and control chicken embryos (-). Fold enrichment of the target region, containing the homeobox binding site versus a negative control region from the coding region of the same gene after anti-flag precipitation is shown. Error bars display standard deviation of the mean.

Beckmann *et al. BMC Developmental Biology* 2011, **11**:35
http://www.biomedcentral.com/1471-213X/11/35

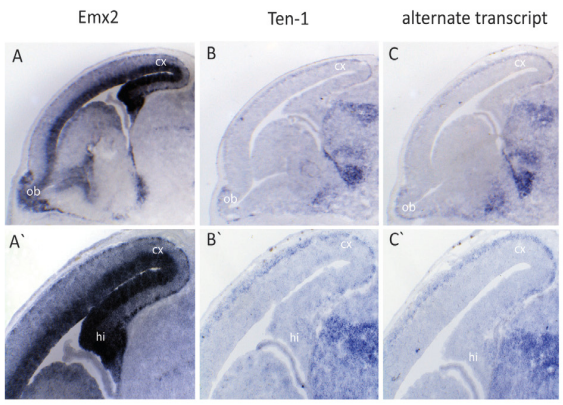


Figure 5 *In-situ* hybridizations for Teneurin-1 and EMX2 expression on sagittal sections at E14.5 mouse embryos. (A; A') Expression of EMX2 in the cortex at higher magnification in A'. (B; B') Expression of total teneurin-1 in an adjacent section. (C; C') Expression of the alternate transcript of teneurin-1 in an adjacent section. Ob, olfactory bulb; hi, hippocampus

promoter activity and is bound directly by EMX2, as shown by gel shift assay and ChIP in chicken embryos. Although one has to take into account that the teneurin-1 expression, especially in later developmental stages, opposes the expression pattern of EMX2, a direct regulation of teneurin-1 expression by EMX2 is likely to occur at earlier stages. First, we and others [26] showed that

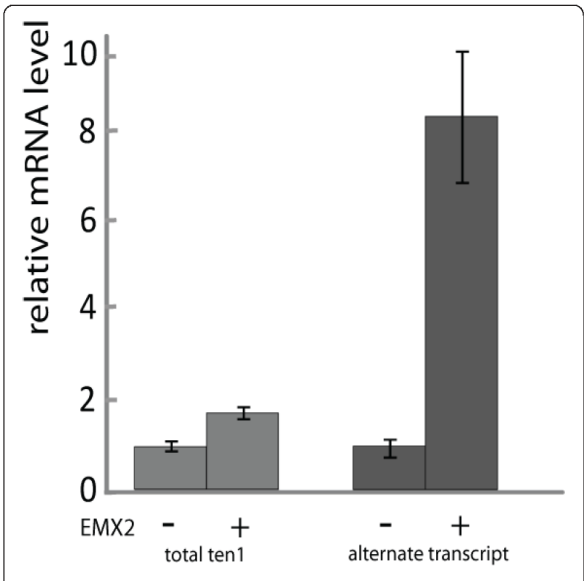


Figure 6 Activation of the alternate transcript by EMX2. RNA was isolated from parental HEK293 cells (-) and HEK293 cells stably expressing EMX2 (+). Teneurin-1 mRNA levels of 3 independent preparations were measured by real-time PCR. The graph shows total teneurin-1 mRNA (total ten1) and mRNA levels under control of the alternate promoter (alternate transcript) relative to GAPDH mRNA values. Values of the parental cells were arbitrarily set to 1. Error bars display standard error of the mean.

total teneurin-1 expression, as well as the expression of the alternate transcript, correlates well with EMX2 expression at E14.5. Secondly, the expression of teneurin-1 is highly dynamic and its patterned expression and the overall expression level collapses in EMX2-deficient mice [26]. Notably, we found teneurin-1 being expressed in the marginal, but not in the ventricular zone of the cortex. This suggests a possible function of EMX2 in post-mitotic neurons via the control of teneurin-1. Based on these findings, it is conceivable that the promoter at the published transcription start is responsible for the basal expression of teneurin-1, whereas the novel promoter is responsible for the graded expression dependent on EMX2. This finding suggests that this promoter region of the teneurin-1 gene is essential in establishing correct patterning of teneurin-1 expression. Although the transcription factors involved in correct patterning and arealization are well known, and their expression patterns are well characterized, very little is known about the downstream mechanisms contributing in establishing proper arealization and pathfinding [reviewed in 45]. A number of reports describe screens to find genes which are differentially expressed within the cortex [24,46,47] or which are potential target genes of differentially expressed transcription factors [11,26,48-50]. Interestingly, in both types of approaches members of the teneurin protein family were revealed as differentially expressed genes, supporting the evidence for a role of teneurin in arealization. Lists of potential target genes of the transcription factors involved in arealization, like Emx2 or Pax6, have also been described in knock-out gene expression studies [26,31,48], but indirect effects on transcription cannot be ruled out and interesting targets need to be validated. In this study, we validated teneurin-1 as the first direct target gene of EMX2 in human. As a transmembrane protein, teneurin-1 is well-suited to convey nuclear signals to the level of cell-cell interactions. However, the molecular mechanisms of how teneurins mediate their proposed function in brain development and patterning of the cortex remain to be elucidated.

Many cases of XLMR have been mapped to Xq25, the locus of the teneurin-1 gene [37-41]. Interestingly, many of these individuals suffer from motor sensory neuropathy [37], and teneurin-1 is predominantly expressed in patterns that relate to anterior sensorimotor areas [26]. Taking into account the regulation of teneurin-1 by EMX2 at the novel promoter, setting up proper arealization of the developing cortex, and the well established functions of teneurins in correct pathfinding and neurite growth, we consider teneurin-1 as a potential target gene for XLMR. When analyzing patient samples, attention should be given to the newly established promoter region, as mutations or deletions in this area could lead to a shift in expression of teneurin-1 early in the developing brain

leading to improper connectivity and consequently to XLMR.

Conclusion

In this work, we show that teneurin-1 expression is regulated by EMX2 at a novel and conserved upstream promoter. We present teneurin-1 as the first direct target gene in humans and characterize the binding site in the newly identified promoter region.

Methods

Rapid amplification of 5' complementary DNA ends (5' RACE)

Total RNA of mouse and chicken brain tissue was purified with QiaShredder and RNA Easy kit (Qiagen, Hombrechtikon, CH) according to manufacturer's instructions. Using these RNA extracts and total human adult normal brain RNA (ams Biotechnology, Oxon, UK) 5'RACEs were performed with the 2nd generation 5'/3'RACE kit (Roche Diagnostic, Mannheim, Germany) according to manufacturer's instructions. Nested PCRs with the primer sequences shown in Table 2 were performed. The bands obtained were purified and cloned into vector pKS$^+$ and sequenced. The sequences were analyzed using the BLAT algorithm on human genome assembly GRCh37, mouse assembly July 2007 and chicken assembly May 2006 (http://genome.ucsc.edu) [51].

Promoter studies

The promoter constructs for human teneurin 1 were amplified from human genomic DNA using the Expand High Fidelity system (Roche) with primer hten1 promoter 1 **XhoI** fw (all primer sequences are given in Table 2) and hten1 promoter 1 **HindIII** rev using the highlighted restriction sites for directional cloning into vector pSEAP2-Basic (Clontech, Mountain View, CA, USA) of the promoter 1 construct. For the hten1 promoter 2 construct, we used hten1 promoter 2 **NheI** fw and hten1 promoter 2 **EcoRI** rev using the highlighted restriction sites for directional cloning into the same vector. The promoter 1 construct contains a sequence of just over 2 kb from nt124097602 to nt124099666 of chromosome x and the promoter 2 construct around 4 kb from nt124336306 to nt124340205 on chromosome x of assembly GRCh37. Analysis of the promoter sequences for potential binding sites was done using the JASPAR database (http://jaspar.cgb.ki.se[52]) and the *ems* matrix. Mutation of the potential homeobox-binding sequence (nt124338584 to nt124338589 on chromosome X) in the promoter 2 was achieved using overlapping PCR with the primer set for hten1 promoter2 and mutated promoter 2 fw and rev. HEK293-EBNA cells were plated at 1 × 10^5 cells per well in six-well plates 18 h before transfection. Cells were transfected in DMEM containing 0.3% FCS with Fugene 6 Transfection reagent (Roche) using 1 μg promoter construct DNA and co-transfected with either empty 1 μg pcDNA3 or flag- and myc-tagged EMX2 in pCMV-Entry (OriGene, Rockville, MD, USA) as indicated in the Result section. Twenty-Four hours after transfection, the medium was collected and SEAP reporter gene activity was measured and normalized for the co-transfected plasmid pGL3, expressing firefly luciferase (Promega, Madison, WI, USA) as previously described [53].

Table 2 Primer sequences

Primer name	Sequence
human RT-PCR	TTAGTGCATGGTCAGGTGAGG
mouse RT-PCR	TCTCCCATCTTCACTCTCATCAG
chicken RT-PCR	GCTGTGTTCTGACTTCATCC
hten PCR1 rev	GTGTCCACATCAGATCCCATCTC
hten nested **XbaI** rev	TAGT**TCTAGA**GCACAGGTGCAGGCATGAGG
mouse PCR1 rev	CTCCAGCTGGTAGCCATGTCG
mouse nested **XbaI** rev	TAGT**TCTAGA**TCTGTGTGGTAGCCGGAGCAC
chicken PCR1 rev	ATGCGCCACCTTCGGTTTCAG
chicken nested **XbaI** rev	TAGT**TCTAGA**GTAGCTGTGCTCCGACACACC
oligo dT anchor primer	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTT
hten1 promotor 1 **XhoI** fw	ACTA**CTCGAG**CAAGACCCATGCTGAAGCT
hten1 promotor 1 **HindIII** rev	ACTA**AAGCTT**CTCTGATTAAGCAAGCTCAGTTTC
hten1 promoter 2 **NheI** fw	ACTA**GCTAGC**CCCCTAGAGTGTTCAGCTCT
hten1 promoter 2 **EcoRI** rev	ACTA**GAATTC**GGGCCACCTCAAAAACACCTCC
mutate promoter 2 fw	CCACCCCTCACCCTCAGGAGAAAGTGGTTAAA
mutate promoter 2 rev	TTTAACCACTTTCTCCTGAGGGTGAGGGGTGG
total ten1 qPCR fw	GCATAGTTCCTGTTTGTCCA
total ten1 qPCR rev	TCTGCACATCTTGAGTAGAC
alt exon qPCR fw	GCTTGGAATAGGAATTACAAGG
alt exon qPCR rev	GAAGTCCTTTAATGCAAGCAG
hGAPDH qPCR fw	GGAGTCAACGGATTTGGTC
hGAPDH qPCR rev	AAACCATGTAGTTGAGGTC
ChIP target region fw	TTCAGCTTCCTCGTTCTTCG
ChIP target region rev	GGTGGTTACAACCGCCTTTT
ChIP negative control fw	AGATTCCTGTGAGCCCTGCT
ChIP negative control rev	TCCAACAACTCATGCAATGG

Real-time Q-PCR

HEK293-ECR cells were transfected as described before with the flag-myc-tagged EMX2 construct in pCMV and cells were selected for stably expressing clones with G418 (Roche) for 2 weeks. Clones were pooled and expression of the construct was tested by Western blot

(data not shown). From these cells and untransfected HEK293-ECR cells RNA was isolated with QiaShredder and the RNA Easy kit (Qiagen) following the manufacturer's protocol. From this preparation, cDNA was generated using the Superscript III (Invitrogen) polymerase and random primers following the standard protocol. Real-time Q-PCR was performed on these samples with teneurin-1 specific primers and normalized to GAPDH values (sequences Table 2) using SYBR QPCR Supermix with ROX (Invitrogen) on an AbiPrism 7000 system. Three independent experiments were performed and the averaged results are shown and p-values were calculated using the one-way ANOVA.

In ovo electroporation

For reporter assay experiments, chicken eggs were incubated in a humidified chamber at 38°gC and DNA constructs were injected into the lumen of the neural tube of stage Hamburger Hamilton (HH) 10-12 embryos. Construct concentrations were: 1 μg/μl lacZ reporter construct (BGZ40; [54]), 1 μg/μl EMX2 expression vector, and 0.2 μg/μl co-injected EGFP in pCMV as positive control of electroporated cells. Embryos were harvested 24 hours after electroporation and processed for β-galactosidase staining. For EMX2 overexpression, 1 μg/μl of Myc/FLAG-tagged EMX2 expression vector and 0.2 μg/μl of pCMV-EGFP construct were co-injected into the lumen of forebrain of stage HH 14 embryos. Positive tissues (n = 20 brains) were collected 72 hours after electoporation and immediately processed for chromatin cross-linking. As negative control, the same amount of unelectroporated tissue was collected and processed for ChIP experiments. Electroporations were performed as described previously using a square wave electroporator [54].

Chromatin immunoprecipitation Assay

Brains were chopped and then cross-linked in 1% Formaldehyde (F8775, Sigma) for 10 minutes at room temperature. Cross-linking was stopped in 125 mM Glycine for 5 minutes and the material was washed three times in ice cold PBS containing EDTA-free Protease Inhibitor Cocktail (Complete, 04693132001, Roche). DNA shearing was performed in lysis buffer (50 mMTris-HCl pH8.0, 10 mM EDTA, 1%SDS, 1 × Protease Inhibitor Cocktail) using the following parameters: 20 cycles of 30 seconds ON/30 seconds OFF (Diagenode bioruptor sonicator, high power setting).

Chromatin immunoprecipitation was performed by using Dynabeads protein G (100.04D, Invitrogen) as described elsewhere [55]. The following antibodies were used: Mouse monoclonal anti-FLAG M2 (F1804, Sigma), Mouse control IgG (AB18413, Abcam).

Electrophoretic Mobility Gel Shift Assay (EMSA)

EMX2 binding to the promoter construct was examined by Electrophoretic Mobility Gel Shift Assay (EMSA) using DIG-labeled double-stranded oligo-nucleotides (5'CAGGAGAAAGTAATTAAAAAA3' or with mutated binding site 5'CAGGAGAAAGTGGTTAAAAAA3', putative binding site underlined). For probe preparation, 5 μg of sense and anti-sense oligo-nucleotides were diluted in 90 μl TE buffer, incubated for 10 min at 95°C and cooled down for 30 min at room temperature for annealing. DIG-labeling of the probes was achieved using the DIG Gel Shift Kit, 2nd Generation (Roche) according to the manufacturer's instructions. For the gel shift assay, nuclear extracts from stably EMX2 expressing HEK293 cells. *in-vitro* translated extracts or nuclear extracts of E12.5 embryo brains containing 20 μg of total protein were incubated with 4 μl of 5 × binding buffer of the Gel Shift Kit, 1 μg double-stranded poly (dIdC) and 0.1 μg poly-L-lysine in a 19 μl reaction mix. For the competition assay, unlabeled wild-type or mutant annealed oligo-nucleotide were added with a 150-fold excess. This mix was incubated for 20 min at room temperature. Afterwards, 1 μl of labeled probe was added and the mix was incubated for another 20 min at 30°C. For supershifts, 1 μl of c-myc antibody (Sigma) was added after 10 min incubation with the labeled probe. Following another 10 min of incubation, the reaction mix was loaded onto a precast 6% DNA retardation gel (Invitrogen, Carlsbad, CA, USA), which was pre-run in 0.5 × TBE for 20 min at 80 V and 4°C. The gel was run for 1.5 h at 80 V and 4°C. After separation, the complexes were blotted on a positively charged nylon membrane in 0.5 × TBE for 45 min at 280 mA and DIG detection was performed as described in the manufacturer's instructions.

In-vitro transcription and translation of EMX2

In-vitro transcription and translation of EMX2 was achieved using the TNT® Coupled Reticulocyte Lysate System (Promega). 25 μl of TNT® rabbit reticulocyte lysate, 2 μl TNT® reaction buffer, 1 μl TNT® T7 RNA polymerase, 0.5 μl of each Amino Acid mixture without Leucine and without Methionine and 1 μg of EMX2-pCMV-Entry were mixed in a 50 μl reaction mix and incubated for 90 min at 30°C, quick frozen in dry ice/ethanol and stored at -80°C until used in EMSA.

In-situ hybridization

In-situ hybridizations on sections were performed as previously described [56]. The following RNA probes were used: For EMX2 we used the entire CDS of EMX2 (NM_010132.2), for total teneurin-1 we used the probe previously published [26] and for the probe specific for

the alternate transcript we used the sequence described in Table 1 (m1_cl_2) plus the first 100 bp of the CDS of mouse teneurin-1 (NM_011855.3).

Abbreviations

XLMR: X-linked mental retardation; hten1: human teneurin-1; EMSA: Electrophoretic Mobility Shift Assays; RACE: rapid amplification of cDNA ends; CRF: corticotrophin-releasing factor; SEAP: secreted embryonic alkaline phosphatase; ChIP: Chromatin Immunoprecipitation

Acknowledgements

We thank Hans-Rudolf Hotz for the help with the UCSC genome browser, Richard P. Tucker for critical reading of the manuscript and Jean-François Spetz for providing the mouse embryos. This work was supported by the Novartis Research Foundation.

Author details

[1]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]University of Basel, Faculty of Science, Basel, Switzerland. [3]Current Address: Department of Radiation Oncology, Division of Radiation and Cancer Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

Authors' contributions

JB prepared all constructs, participated in the 5'RACE, performed sequence alignments, EMSA and QPCR and prepared the first draft of the manuscript. AV performed the chicken electroporation and the ChIP experiments as well as the mouse *in-situ* hybridization. JF performed the SEAP assay and validated EMX2 expression. DKB performed the 5'RACE in chicken and mouse. FMR participated in the planning and discussion of the experiments. RCE participated in the planning and discussion of the experiments, writing the paper and preparation of the figures. All authors read and approved the final manuscript.

Received: 27 September 2010 Accepted: 8 June 2011
Published: 8 June 2011

References

1. Tucker RP, Kenzelmann D, Trzebiatowska A, Chiquet-Ehrismann R: **Teneurins: Transmembrane proteins with fundamental roles in development.** *The International Journal of Biochemistry & Cell Biology* 2007, **39**:292-297.
2. Baumgartner S, Chiquet-Ehrismann R: **Tena, a *Drosophila* gene related to tenascin, shows selective transcript localization.** *Mechanisms of Development* 1993, 165-176.
3. Minet AD, Chiquet-Ehrismann R: **Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli*.** *Gene* 2000, **257**:87-97.
4. Lovejoy DA, Rotzinger S, Barsyte-Lovejoy D: **Evolution of Complementary Peptide Systems.** *Annals of the New York Academy of Sciences* 2009, **1163**:215-220.
5. Kenzelmann D, Chiquet-Ehrismann R, Leachman NT, Tucker RP: **Teneurin-1 is expressed in interconnected regions of the developing brain and is processed *in vivo*.** *BMC Developmental Biology* 2008, **8**:30.
6. Nunes SM, Ferralli J, Choi K, Brown-Luedi M, Minet AD, Chiquet-Ehrismann R: **The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix.** *Experimental Cell Research* 2005, **305**:122-32.
7. Bagutti C, Forro G, Ferralli J, Rubin B, Chiquet-Ehrismann R: **The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription.** *Journal of Cell Science* 2003, **116**:2957-66.
8. Drabikowski K, Trzebiatowska A, Chiquet-Ehrismann R: **ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*.** *Developmental Biology* 2005, **282**:27-38.
9. Trzebiatowska A, Topf U, Sauder U, Drabikowski K, Chiquet-Ehrismann R: ***Caenorhabditis elegans* teneurin, ten-1, is required for gonadal and pharyngeal basement membrane integrity and acts redundantly with integrin ina-1 and dystroglycan dgn-1.** *Molecular Biology of the Cell* 2008, **19**:3898-908.
10. Mörck C, Vivekanand V, Jafari G, Pilon M: ***C. elegans* ten-1 is synthetic lethal with mutations in cytoskeleton regulators, and enhances many axon guidance defective mutants.** *BMC Developmental Biology* 2010, **10**:55.
11. Baumgartner S, Martin D, Hagios C, Chiquet-Ehrismann R: **Ten(m), a *Drosophila* gene related to Tenascin, is a new pair-rule gene.** *EMBO Journal* 1994, **13**:3728-3740.
12. Levine A, Bashan-Ahrend A, Budai-Hadrian O, Gartenberg D, Menasherow S, Wides R: ***odd Oz*: A novel Drosophila pair rule gene.** *Cell* 1994, **77**:587-598.
13. Rakovitsky N, Buganim Y, Swissa T, Kinel-Tahan Y, Brenner S, Cohen MA, Levine A, Wides R: ***Drosophila* Ten-a is a maternal pair-rule and patterning gene.** *Mechanisms of Development* 2007, **124**:911-924.
14. Kinel-Tahan Y, Weiss H, Dgany O, Levine A, Wides R: ***Drosophila* odz gene is required for multiple cell types in the compound retina.** *Developmental Dynamics* 2007, **236**:2541-2554.
15. Ben-Zur T, Feige E, Motro B, Wides R: **The Mammalian Odz Gene Family: Homologs of a *Drosophila* Pair-Rule Gene with Expression Implying Distinct yet Overlapping Developmental Roles.** *Developmental Biology* 2000, **217**:107-120.
16. Oohashi T, Zhou X-H, Feng K, Richter B, Morgelin M, Perez MT, Su W-D, Chiquet-Ehrismann R, Rauch U, Fassler R: **Mouse Ten-m/Odz Is a New Family of Dimeric Type II Transmembrane Proteins Expressed in Many Tissues.** *Journal of Cell Biology* 1999, **145**:563-577.
17. Zhou X-H, Brandau O, Feng K, Oohashi T, Ninomiya Y, Rauch U, Fässler R: **The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain.** *Gene Expression Patterns* 2003, **3**:397-405.
18. Rubin BP, Tucker RP, Brown-Luedi M, Martin D, Chiquet-Ehrismann R: **Teneurin 2 is expressed by the neurons of the thalamofugal visual system *in situ* and promotes homophilic cell-cell adhesion *in vitro*.** *Development* 2002, **129**:4697-705.
19. Tucker RP, Martin D, Kos R, Chiquet-Ehrismann R: **The expression of teneurin-4 in the avian embryo.** *Mechanisms of Development* 2000, **98**:187-91.
20. Tucker RP, Chiquet-Ehrismann R, Chevron MP, Martin D, Hall RJ, Rubin BP: **Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced *in vitro* and *in situ* by FGF8.** *Developmental Dynamics* 2001, **220**:27-39.
21. Otaki JM, Firestein S: **Neurestin: Putative Transmembrane Molecule Implicated in Neuronal Development.** *Developmental Biology* 1999, **212**:165-181.
22. Mieda M, Kikuchi Y, Hirate Y, Aoki M, Okamoto H: **Compartmentalized expression of zebrafish ten-m3 and ten-m4, homologues of the Drosophila tenm/odd Oz gene, in the central nervous system.** *Mechanisms of Development* 1999, **87**:223-227.
23. Feng K, Zhou X-H, Oohashi T, Märgelin M, Lustig A, Hirakawa S, Ninomiya Y, Engel Jr, Rauch U, Fässler R: **All Four Members of the Ten-m/Odz Family of Transmembrane Proteins Form Dimers.** *Journal of Biological Chemistry* 2002, **277**:26128-26135.
24. Leamey CA, Glendining KA, Kreiman G, Kang N-D, Wang KH, Fassler R, Sawatari A, Tonegawa S, Sur M: **Differential Gene Expression between Sensory Neocortical Areas: Potential Roles for Ten_m3 and Bcl6 in Patterning Visual and Somatosensory Pathways.** *Cerebral Cortex* 2008, **18**:53-66.
25. Leamey CA, Merlin S, Lattouf P, Sawatari A, Zhou X, Demel N, Glendining KA, Oohashi T, Sur M, Fässler R: **Ten_m3 Regulates Eye-Specific Patterning in the Mammalian Visual Pathway and Is Required for Binocular Vision.** *PLoS Biology* 2007, **5**:e241.
26. Li H, Bishop KM, O'Leary DD: **Potential target genes of EMX2 include Odz/Ten-M and other gene families with implications for cortical patterning.** *Molecular and Cellular Neuroscience* 2006, **33**:136-49.
27. Sansom SN, Livesey FJ: **Gradients in the brain: the control of the development of form and function in the cerebral cortex.** *Cold Spring Harb Perspect Biol* 2009, **1**:a002519.
28. Dalton D, Chadwick R, McGinnis W: **Expression and embryonic function of empty spiracles: a *Drosophila* homeo box gene with two patterning functions on the anterior-posterior axis of the embryo.** *Genes & Development* 1989, **3**:1940-1956.
29. Simeone A, Gulisano M, Acampora D, Stornaiuolo A, Rambaldi M, Boncinelli E: **Two vertebrate homeobox genes related to the *Drosophila***

empty spiracles gene are expressed in the embryonic cerebral cortex. *EMBO Journal* 1992, 11:2541-2550.
30. Fukuchi-Shimogori T, Grove EA: **Emx2 patterns the neocortex by regulating FGF positional signaling.** *Nature Neuroscience* 2003, **6**:825-831.
31. Bishop KM, Goudreau G, O'Leary DDM: **Regulation of Area Identity in the Mammalian Neocortex by Emx2 and Pax6.** *Science* 2000, **288**:344-349.
32. Gulisano M, Broccoli V, Pardini C, Boncinelli E: **Emx1 and Emx2 show different patterns of expression during proliferation and differentiation of the developing cerebral cortex in the mouse.** *European Journal of Neuroscience* 1996, **8**:1037-1050.
33. Mallamaci A, Iannone R, Briata P, Pintonello L, Mercurio S, Boncinelli E, Corte G: **EMX2 protein in the developing mouse brain and olfactory area.** *Mechanisms of Development* 1998, **77**:165-172.
34. O'Leary DDM, Chou S-J, Sahara S: **Area Patterning of the Mammalian Cortex.** *Neuron* 2007, **56**:252-269.
35. Otaki JM, Firestein S: **Segregated expression of neurestin in the developing olfactory bulb.** *Neuroreport* 1999, **10**:2677-2680.
36. Minet AD, Rubin BP, Tucker RP, Baumgartner S, Chiquet-Ehrismann R: **Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain.** *Journal of Cell Science* 1999, **112(Pt 12)**:2019-32.
37. Malmgren H, Sundvall M, Dahl N, Gustavson KH, Anneren G, Wadelius C, Steenbondeson ML, Pettersson U: **Linkage maping of a severe X-linked mental-retardation syndrome.** *American Journal of Human Genetics* 1993, **52**:1046-1052.
38. Gustavson KH, Anneren G, Malmgren H, Dahl N, Ljunggren CG, Backman H: **New X-linked syndrome with severe mental retardation, severely impaired vision, severe hearing defect, epileptic seizures, spasticity, restricted joint mobility and early death.** *American Journal of Medical Genetics* 1993, **45**:654-658.
39. Cabezas DA, Slaugh R, Abidi F, Arena JF, Stevenson RE, Schwartz CE, Lubs HA: **A new X linked mental retardation (XLMR) syndrome with short stature, small testes, muscle wasting, and tremor localises to Xq24-q25.** *Journal of Medical Genetics* 2000, **37**:663-668.
40. Martínez F, Martínez-Garay I, Oltra S, Moltó MD, Orellana C, Monfort S, Prieto F, Tejada I: **Localization of MRX82: A new nonsyndromic X-linked mental retardation locus to Xq24-q25 in a Basque family.** *American Journal of Medical Genetics* 2004, , **131A**: 174-178.
41. Vitale E, Specchia C, Devoto M, Angius A, Rong S, Rocchi M, Schwalb M, Demelas L, Paglietti D, Manca S, *et al*: **Novel X-linked mental retardation syndrome with short stature maps to Xq24.** *American Journal of Medical Genetics* 2001, **103**:1-8.
42. Iler N, Rowitch DH, Echelard Y, McMahon AP, Abate-Shen C: **A single homeodomain binding site restricts spatial expression of Wnt-1 in the developing brain.** *Mechanisms of Development* 1995, **53**:87-96.
43. Shepard S, McCreary M, Fedorov A: **The Peculiarities of Large Intron Splicing in Animals.** *PLoS ONE* 2009, **4**:e7853.
44. Singh J, Padgett RA: **Rates of in situ transcription and splicing in large human genes.** *Nat Struct Mol Biol* 2009, **16**:1128-1133.
45. O'Leary DDM, Sahara S: **Genetic regulation of arealization of the neocortex.** *Current Opinion in Neurobiology* 2008, **18**:90-100.
46. Funatsu N, Inoue T, Nakamura S: **Gene Expression Analysis of the Late Embryonic Mouse Cerebral Cortex Using DNA Microarray: Identification of Several Region- and Layer-specific Genes.** *Cereb Cortex* 2004, **14**:1031-1044.
47. Mühlfriedel S, Kirsch F, Gruss P, Chowdhury K, Stoykova A: **Novel genes differentially expressed in cortical regions during late neurogenesis.** *European Journal of Neuroscience* 2007, **26**:33-50.
48. Gangemi RMR, Daga A, Muzio L, Marubbi D, Cocozza S, Perera M, Verardo S, Bordo D, Griffero F, Capra MC, *et al*: **Effects of Emx2 inactivation on the gene expression profile of neural precursors.** *European Journal of Neuroscience* 2006, **23**:325-334.
49. Holm PC, Mader MT, Haubst N, Wizenmann A, Sigvardsson M, Götz M: **Loss- and gain-of-function analyses reveal targets of Pax6 in the developing mouse telencephalon.** *Molecular and Cellular Neuroscience* 2007, **34**:99-119.
50. Numayama-Tsuruta K, Arai Y, Takahashi M, Sasaki-Hoshino M, Funatsu N, Nakamura S, Osumi N: **Downstream genes of Pax6 revealed by comprehensive transcriptome profiling in the developing rat hindbrain.** *BMC Developmental Biology* 2005, **10**:6.
51. Kent WJ: **BLAT - The BLAST-Like Alignment Tool.** *Genome Research* 2002, **12**:656-664.
52. Bryne JC, Valen E, Tang M-HE, Marstrand T, Winther O, Piedade I da, Krogh A, Lenhard B, Sandelin A: **JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.** *Nucl Acids Res* 2008, **36**:D102-106.
53. Degen M, Goulet S, Ferralli J, Roth M, Tamm M, Chiquet-Ehrismann R: **Opposite effect of fluticasone and salmeterol on fibronectin and tenascin-C expression in primary human lung fibroblasts.** *Clinical & Experimental Allergy* 2009, **39**:688-699.
54. Itasaki N, Bel-Vialar S, Krumlauf R: **'Shocking' developments in chick embryology: electroporation and *in ovo* gene expression.** *Nature Cell Biology* 1999, 1:E203-E207.
55. Vitobello A, Ferretti E, Lampe X, Vilain N, Ducret S, Ori M, Spetz J-F, Selleri L, Rijli M Filippo: **Hox and Pbx Factors Control Retinoic Acid Synthesis during Hindbrain Segmentation.** *Developmental Cell* **20**:469-482.
56. Minoux M, Antonarakis GS, Kmita M, Duboule D, Rijli FM: **Rostral and caudal pharyngeal arches share a common neural crest ground pattern.** *Development* 2009, **136**:637-645.

doi:10.1186/1471-213X-11-35
Cite this article as: Beckmann *et al.*: Human teneurin-1 is a direct target of the homeobox transcription factor EMX2 at a novel alternate promoter. *BMC Developmental Biology* 2011 11:35.

4.3. Teneurin homophilic interaction is mediated by their NHL repeat domain and inhibits neurite outgrowth

Jan Beckmann, Rajib Schubert, Ruth Chiquet-Ehrismann and Daniel Müller

Manuscript in preparation

My contribution to this paper:

For this study, I cloned every construct and established all cell lines used in the manuscript. Subsequently, I tested all cell lines for correct surface expression of the different constructs and performed the FACS sorting. Furthermore, I performed the co-culture assay with primary hippocampal cells and the neurite outgrowth assay of Nb2a cells. Finally, I prepared the first draft of the manuscript and the final manuscript with input of all co-authors.

Teneurin homophilic interaction is mediated by their NHL repeat domain and inhibits neurite outgrowth

Jan Beckmann[1,2,4], Rajib Schubert[3,4], Ruth Chiquet-Ehrismann[1,2,**] and Daniel J. Müller[3,*]

[1]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

[2]University of Basel, Faculty of Science, Basel, Switzerland

[3]Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland

[4]These authors contributed equally to this work

[*]Correspondence: daniel.mueller@bsse.ethz.ch T:+41 61 387 33 07

**Correspondence: ruth.chiquet@fmi.ch T: +41 61 697 24 94 F: +41 61 697 39 76

Results

Summary

The teneurins are highly conserved transmembrane proteins and function as axon guidance and/or target selection molecules in the developing brain. The molecular mechanisms and protein domains mediating these functions are unknown. Using single cell force spectroscopy we show that teneurins mediate increased adhesion forces in a homophilic, but not heterophilic manner between two contacting cells. Furthermore, we observed an increase in adhesion force over time that is depended on the presence of the intracellular domain. Using a combinatorial approach of domain deletions as well as domain exchanges between the large multi-domain extracellular parts of teneurin-1 and teneurin-2, we found that the NHL repeat domain is responsible for discriminating homophilic versus heterophilic interactions. Finally, we tested the effect of homophilic versus heterophilic interaction on neuronal cells and found that neurite outgrowth of Nb2a cells was slowed down by homophilic interaction.

Results

Introduction

The large type II transmembrane proteins of the teneurin family are highly conserved from invertebrates to vertebrates, regarding their domain architecture and amino acid sequence (reviewed in Tucker et al. 2007; Tucker et al. 2011). In vertebrates, four paralogs called teneurin-1 to -4 exist. Extensive localization studies of teneurin expression in vertebrates revealed distinct non-overlapping patterns of expression in specific subpopulations of neurons for the different family members (reviewed in Kenzelmann et al. 2007; Young and Leamey 2009).

In vivo studies of different teneurin-deficient model organisms revealed the importance of teneurins for axon guidance and target recognition in the developing nervous system. In *Caenorhabditis elegans*, RNAi knockdown or teneurin mutants of its single gene lead to multiple phenotypes including failures in axon guidance and neural pathfinding (Drabikowski et al. 2005; Mörck et al. 2010). In *Drosophila melanogaster* a recent study showed that knock-out of *ten-m*, one of the two fly teneurin genes, as well as epithelial overexpression of the ten-m protein, leads to motor neuron guidance defects (Zheng et al. 2012). Furthermore, both of the two *Drosophila* teneurin proteins are shown to be involved in synapse organization and target selection based on homophilic interactions (Hong et al. 2012; Mosca et al. 2012). In vertebrates, teneurin-3 deficient mice show changes in the connections of the ipsilateral retinal inputs, resulting in binocular mismatch and major deficits in the performance of visually mediated tasks (Leamey et al. 2007). Furthermore, teneurin-2 was shown to be a transsynaptic binding partner for Latrophilin 1 (LPH1) and this binding induces LPH1 mediated calcium signaling (Silva et al. 2011).

All these reports clearly identify teneurin proteins as important and conserved target finding and axon guidance molecules for specific subsets of neurons. Projecting neurons find their way to their respective target site by integrating various guidance stimuli with their growth cone along the way to the target region. They often follow tracks of attractive molecules either deposited in the extracellular matrix or still bound to the membrane of other neurons or glia cells along these paths (reviewed in Dickson 2002; Zipursky and Sanes 2010). Previous studies in chicken and *Drosophila* show low levels of teneurin expressed along the projections of neurons (Rubin et al. 2002; Hong, Mosca et al. 2012). Several *in vitro* and cell culture experiments suggested homophilic interactions between teneurins to underlie the

reported *in vivo* functions (Oohashi et al. 1999; Rubin, Tucker et al. 2002; Bagutti et al. 2003).

In the present study we quantified the adhesion forces between teneurin expressing cells using an atomic force microscope (AFM) as a single cell force spectroscope (SCFS; (Helenius et al. 2008)), and identify the intracellular as well as extracellular domains of teneurins responsible for homophilic interactions and generation of adhesion strength. Finally, we provide evidence for an inhibitory signal for neurite outgrowth by homophilic teneurin interactions.

Results

Primary hippocampal neurons grow neurites on teneurin-expressing cells

The general teneurin structure is presented in Fig. 1A. The N-terminal intracellular domain (ICD; I) is followed by a single span transmembrane domain, and a large extracellular domain consisting of eight tenascin- type EGF-like repeats (E), a cystine-rich domain with five NHL repeats (N) and a YD-repeat domain (Y). To test if teneurin can act as a membrane-bound attractive molecule for primary neurons, we set up a co-culture assay of teneurin-2 expressing non-neuronal HEK293 cells expressing either N-terminal GFP-tagged full-length teneurin-2 (I2E2N2Y2; for domain denomination code cf. Fig. 1A,B) or a GFP-tagged construct lacking the ECD (I2E0N0Y0) and dissociated primary hippocampal cells using the method described in Biederer and Scheiffele, 2007. Analyzing co-localization of synapsin1 (Syn1) as an axonal marker and GFP for the HEK293 cells we observed that the primary neurons grow preferentially over the HEK293 cells expressing full-length teneurin-2 (Fig. 1C, left panel, D) avoiding the HEK293 cells expressing the truncated construct (Fig. 1C, right panel, D). Teneurins were shown to function as target recognition molecules, mediated by homophilic interactions (Hong et al., 2012; Mosca et al., 2012). Further, dissociated primary hippocampal cells express teneurin-2 (Silva et al., 2011) and show preferences for a teneurin-2 containing growth substrate in our assay. Thus, we decided to investigate the mechanism of teneurin interactions in more detail.

Teneurins interact in a homophilic, but not in a heterophilic manner

To address the question whether teneurins can act as homophilic and/or heterophilic adhesion proteins under well controlled conditions we created HEK293 cells expressing full-length teneurin-1 (I1E1N1Y1) or teneurin-2 (I2E2N2Y2). Using SCFS as depicted in Fig. 2A, we tested in a quantifiable manner, if teneurins interact on cell surfaces of two contacting cells. Measuring the maximal adhesion forces of cells both expressing either full-length teneurin-1 (I1E1N1Y1 vs I1E1N1Y1) or teneurin-2 (I2E2N2Y2 vs I2E2N2Y2) resulted in similarly high adhesion forces, which were significantly increased compared to the adhesion forces of control cells lacking the teneurin extracellular domain (ECD; I2E0N0Y0 vs I2E0N0Y0); Fig. 2B,C). As can be seen in Fig. 2B, increased adhesion forces started to be measurable after 20 seconds of cell contact and increased with time reaching an adhesion force of about 2nN after 60 seconds and 4nN after 120 seconds of contact. To exclude that teneurins interact

with an endogenous ligand on the HEK293 cells we analyzed the adhesion forces between cells expressing either one of the full-length teneurins with cells lacking the ECD (I2E0N0Y0 vs I2E2N2Y2 and I2E0N0Y0 vs I1E1N1Y1) which resulted in lo,w adhesion forces (Fig.2C). Furthermore, cells expressing teneurin-1 on one cell and teneurin-2 on the second cell (I1E1N1Y1 vs I2E2N2Y2) also resulted in low adhesion forces (Fig. 2C), proving that teneurins do not interact heterophilically in the cellular context. The recorded differences in the observed adhesion forces is unlikely to be a consequence of differences in cell mechanical properties, since we do not observe a correlation between contact stiffness and the maximum adhesion forces (Fig. 2D). To ensure similar expression levels of the constructs the cells were FACS sorted with the same gating (Fig. S1A,B) and surface expression of the teneurin constructs was confirmed with antibody staining against the EGF domains of teneurin-1 or -2 of non-permeabilized cells in comparison to the respective GFP- or RFP-tag (Fig.2E and S1A,B).

Teneurin mediated adhesion force is dependent on the intracellular domain and independent of the YD-repeats

The YD repeat domain is unique for teneurins among all eukaryotic proteins and makes up for almost half the size of the protein. To test whether the YD repeat domain is involved in teneurin-specific adhesion we analyzed the adhesion forces of cells expressing constructs lacking the complete YD repeat domain (I2E2N2Y0 vs I2E2N2Y0; and I1E1N1Y0 vs I1E1N1Y0 not shown) and found the same high adhesion forces as for full-length teneurin expressing cells (Fig. 3A iii and iv). This finding indicates that the large YD repeat domain is dispensable for the increased adhesion forces mediated by teneurin homophilic interactions. A previous report showed that the ICD of teneurin-1 interacts with the cytoskeletal adaptor protein CAP/ponsin, demonstrating a possible link of teneurins to the cytoskeleton (Nunes et al. 2005). We tested the function of the teneurin ICD domain in mediating the adhesion forces using SCFS. Cells expressing a teneurin-2 construct lacking the ICD (I0E2N2Y2 vs I0E2N2Y2) only show low adhesion forces in the range of control cells lacking the ECD (I2E0N0Y0 vs I2E0N0Y0; Fig. 3A ii). Thus, for strong adhesion between teneurin expressing cells both the ICD as well as parts of the ECD are required, implicating a trans interaction between the proteins on the outside of the cell as well as a possible clustering of the teneurins within the cell membrane by their interaction with the cytoskeleton. As can be seen in Fig. 3B, the action of the ICD is not teneurin-type specific and cells expressing a swap construct

containing the ICD of teneurin-1 and the extracellular part of teneurin-2 (I1E2N2Y2 vs I1E2N2Y2) show the same increase in adhesion forces mediated by homophilic interactions as the original full-length teneurin-2 (I2E2N2Y2 vs I2E2N2Y2). These adhesion forces are also increased to the same level, when cells expressing the ICD swap construct were in contact with cells expressing full-length teneurin-2 (I1E2N2Y2 vs I2E2N2Y2), but not with full-length teneurin-1 (I1E2N2Y1 vs I1E1N1Y1).

Homophilic interaction is mediated by the NHL propeller domain

Since deleting the YD repeats and swapping the ICD domain has no influence in the increased adhesion forces measured, we decided to test which of the remaining domains, the EGF-like repeats or the NHL repeat domain, is responsible for discriminating homophilic versus heterophilic interactions. For this purpose we generated swap constructs of the EGF-like repeats and the NHL repeat domain (cf. Fig. 1B) and tested these constructs for their capability to mediate increased adhesion forces upon homophilic interactions (Fig. 3C). We observed that only cells expressing constructs containing the same type of NHL repeat domain showed increased adhesion forces (compare I1E1N2Y0 vs I1E1N2Y0 and I1E1N1Y0 vs I1E1N2Y0; Fig. 3C). The source of the EGF-like repeats and the ICD do not influence the interaction and it does not matter whether they come from the same teneurin member or the other one (compare I1E1N2Y0 vs I1E1N2Y0 and I2E2N2Y0 vs I1E1N2Y0; Fig. 3C and Fig. S2). This clearly indicates that the homophilic versus heterophilic discrimination between teneurin-1 and -2 is mediated by their respective NHL repeat domain and this is consistently observed for the NHL repeat domain of teneurin-1 as well as teneurin-2 (Fig. 3C; S2).

Teneurin homophilic interaction via the NHL propeller domain inhibits Nb2a neurite outgrowth

To test whether homophilic interactions of teneurin proteins have a function in the outgrowth of neurons, we set up a neurite outgrowth assay on recombinantly expressed and purified teneurin-2 ECD protein. Purity of the protein and coating efficiency of coverslips with the purified protein is shown in Fig. S3A,B. We compared the length of neurites from Nb2a cells expressing different teneurin constructs after differentiation in serum free media on purified teneurin-2 ECD or on poly-lysine as a control substrate. Cells transiently transfected with full-length teneurin-2 (I2E2N2Y2) showed fewer neurites when plated on purified teneurin-2 ECD, compared to neighboring untransfected cells with normal neurites

(Fig. 4A, top left panel). In contrast, cells transiently expressing full-length teneurin-1 (I1E1N1Y1) show normal neurite lengths on teneurin-2 ECD (Fig. 4A, top right panel). No difference is observed between untransfected cells or cells expressing teneurin-1 or teneurin-2 when plated on a control substrate. In order to quantify this observation and to identify the teneurin-2 domains involved in mediating the effect we generated stable cell lines and quantified neurite length of cells expressing different teneurin constructs (Fig. 4B,C). Using cells stably expressing the full-length form of teneurin-2 (I2E2N2Y2) we confirmed the results observed by transient transfection and these cells showed significantly shorter neurites when grown on purified teneurin-2 ECD than control cells lacking the complete ECD (I2N0E0Y0; Fig. 4B, left panel). Interestingly, cells expressing teneurin-2 lacking the ICD (I0E2N2Y2; Fig. 4B) showed no reduction in neurite outgrowth, even though the interacting ECD is present in these cells. No difference in neurite outgrowth is observed when the same cells were seeded on poly-lysine, where all cells grew neurites to the same extent (Fig. 4B, right panel). This indicates that the homophilic interactions of teneurin-2 expressing cells with a teneurin-2 substratum are inhibitory to neurite outgrowth and the effect is only seen when both, the extracellular and the intracellular domains are present. We next tested the functional influence of the homophilic interaction of the NHL propeller domains on neurite outgrowth. Cells expressing constructs containing the propeller domain of teneurin-2 (I2E2N2Y2 and I2E2N2Y0) showed shorter neurites when plated on the purified teneurin-2 ECD, whereas cells expressing swap constructs with the propeller domain of teneurin-1 (I2E2N1Y0) show significantly longer neurites (Fig. 4C, left panel). This clearly indicates that the homophilic interaction mediated by the NHL propeller domain is responsible for the reduced growth rate of neurites. All cells showed the same rate of neurite outgrowth when plated on poly-lysine as a control substrate (Fig. 4C, right panel). Representative images used for the quantification are shown in Fig. S3C,D.

Results

Discussion

Recently it was shown that *Drosophila* teneurins are required for proper target finding of subsets of neurons in the olfactory system. In this study it is shown that projecting olfactory receptor neurons (ORN) with high levels of ten-m protein, one of two *Drosophila* teneurins, make connections only with projecting neurons (PN), which themselves express high ten-m levels. The same is true for ORNs and PNs expressing the second *Drosophila* teneurin, ten-a (Hong, Mosca et al. 2012). These results clearly prove that homophilic interactions between teneurin expressing neurons mediate proper target recognition. However, the mechanism of recognition and the domains responsible have not been analyzed to date.

Different functions have been attributed to the teneurin ICD, ranging from transcriptional mediator after release from the membrane (Bagutti et al. 2003; Nunes et al. 2005) to its interaction with the cyoskeletal adapter protein CAP/ponsin (Nunes et al. 2005). Here we demonstrate that increased adhesion forces between two contacting cells, mediated by homophilic interactions of teneurins, are dependent on the ICD. Furthermore, the ICD of teneurin-1 and -2 are interchangeable and thus seem to have a common function in mediating these forces, indicating a general link to the cytoskeleton for the teneurin proteins. Since longer contact times between two cells lead to increased adhesion forces we envisage that local clustering of teneurins mediated by the cytoskeleton leads to increased local concentration of teneurins. This will increase their avidity through multiple bond interactions thereby increasing the interaction strength between cells. This has been observed also for other cell-cell or cell-extracellular matrix adhesion proteins such as SynCAM1, nectin, or integrins (Fogel et al. 2011; Kurita et al. 2011; Margadant et al. 2011). Previous reports of *in vitro* data suggested that the homophilic interactions of teneurin are mediated by domains C-terminal of the EGF-like repeats (Feng et al. 2002). Using electron microscopy of recombinantly expressed teneurin ECDs Feng et al. (2002) observed three large globular subdomains connected to the rod like EGF-like repeats. The NHL repeat domain is located C-terminal of the EGF-repeats followed by the YD-repeats. NHL repeats are known to form beta-propellers (Edwards et al. 2003) and in teneurins the five NHL repeats most likely form a five-bladed beta propeller domain which most likely represent one of these observed subdomains of the teneurin ECD. We were able to clearly pinpoint the NHL repeat domains to mediate the homophilic interaction between cell surface expressed

teneurins. This is reminiscent of other cell surface receptor interactions. The interaction between the cell surface receptor pairs of semaphorins and plexins are also mediated by propeller–propeller interactions (Janssen et al. 2010; Liu et al. 2010). Whereas in that case the propellers might influence the respective pairing of members of two different protein families, in the case of teneurins only identical propeller domains interact with each other and ensure contacts between cells that express the same type of teneurin only.

These homophilic teneurin interactions have functional consequences on neurons as neurite outgrowth of Nb2a cells is attenuated. This effect is also dependent on the presence of the ICD. In this paradigm, clustering of teneurins on a growth cone adhering to a substratum highly enriched in the same type of teneurin might either lead to strong enough adhesion that further outgrowth is inhibited, or the adhesion triggers an intracellular signaling cascade leading to growth cone arrest. This *in vitro* inhibition of neurite outgrowth is reminiscent of the wiring and target finding that was observed to depend on teneurins *in vivo,* where axons extend until they find partners that express high levels of the same type of teneurin to connect with.

Experimental procedures

Cloning of constructs

Full-length chicken teneurin-1 (I1E1N1Y1; NM_204862) fused to an N-terminal RFP-tag and full-length chicken teneurin-2 (I2E2N2Y2; NM_204097) fused to an N-terminal GFP-tag were cloned into pcDNA3.1 vector. All subsequent swap and deletion constructs were cloned into the same vector. Swap constructs were generated by overlapping PCR using a 14bp overhang. Swap and deletion constructs (Fig. 1B) were cloned using the following domain boundaries in basepairs: ICD (I) 1-852, EGF-like repeats (E) 853-2310, NHL repeats (N) 2311-4428 and YD-repeats (Y) 4429-8118 for teneurin-1. For teneurin-2 the boundaries are I 1-388, E 389-2613, N 2614-4719 and Y 4720-8409. All constructs were sequence verified.

Cell culture

HEK293-ECR (Invitrogen) and Nb2a neuroblastoma cells (ATCC CCL131) were maintained in DMEM media with 10% (vol) FCS, 100 mg/ml penicillin and 100 mg/ml streptomycin (growth medium; all Invitrogen). Stable cell lines were established by G418 selection and FACS sorted for homogenous GFP- or RFP-fusion protein expression with the same gating (Fig. S1). Surface staining of all constructs was confirmed by staining of the non-permeabilized cells with polyclonal antibodies against teneurin-1 or teneurin-2 EGFs described in (Kenzelmann et al. 2008). Z-stacks were acquired on a spinning disk system (Zeiss AxioImager M1 + Yokogawa CSU-22 scanhead) with 100x magnification and the pictures were deconvolved using the Huygens Remote Manager (Ponti et al. 2007).

Cell adhesion measurements using SCFS

One day prior to SCFS, the growth medium was replaced with Liebowitz medium L-15 and supplemented with 1% (vol) FCS. For SCFS, cells were washed with phosphate buffered saline (PBS), detached using 1% (vol) ethylenediaminetetraacetic acid (EDTA, Sigma, Germany) in PBS, pelleted, and resuspended into 1% (vol) serum Liebowitz medium (SCFS measurement medium). Cell-cell adhesion measurements were conducted with an AFM (Nanowizard I, JPK Instruments) mounted on an inverted fluorescence microscope (Zeiss Axiovert 200, equipped with 20x and 40x lenses). The setup was extended with a JPK CellHesion® module to increase the cell-cell separation distance to 100 µm and used in

closed height feedback mode (Puech et al. 2006). Phase-contrast and fluorescence imaging were used to monitor cellular morphology during adhesion measurements. Tip-less AFM cantilevers (NPO-010, Veeco Instruments, nominal spring constant k = 60 mN/m) were calibrated using the thermal noise method (Butt and Jaschke 1995). Each cantilever was calibrated three times to rule out possible errors. Spring constants were consistent within 10% of their nominal value. Plasma-activated cantilevers were incubated with 2.5 mg/ml concanavalin A (ConA, Sigma) overnight at 4°C and carefully rinsed in PBS before use. Plasma-activated petri dishes (WPI, Inc) were prepared to obtain adhesive Petri dishes for HEK cells to adhere. Before SCFS, Petri dishes were gently rinsed with the cell culture medium used to perform the adhesion tests (Liebowitz L-15 medium supplemented with 1% (vol) FCS, 100 mg/ml penicillin and 0.1 mg/ml streptomycin). Diluted cell suspensions were then seeded onto the substrate and incubated over night at 37°C. For SCFS, cells were selected using phase-contrast microscopy and fluorescence microscopy to characterize cell morphology and to ensure consistent teneurin expression levels, respectively (Figure S1c). Once the cell has been selected, the ConA-coated cantilever was gently pressed onto the cell applying a force of ≈1 nN for ≈3 s. After this, the cell attached to the cantilever was withdrawn from the substrate for 2–10 min to allow the cell to enhance this attachment. This ‘probe-cell’ was then moved above a 'target-cell’ that was firmly attached to the adhesive ConA-coated part of the substrate. Cell adhesion experiments between probe-cell and target-cell (see Figure 2a) were performed applying a contact force of 1 nN, contact times ranging from 5–120 s, and 5 µm/s approach and retract velocities. The contact time was varied randomly for a given cell-cell couple to prevent systematic bias or history effect. Each force-distance (F-D) curve characterizing the adhesion between probe-cell and target-cell was repeated up to many times depending on the contact time: 5 s contact time, 5 repetitions; 20 s contact time, 3 repetitions; 60 s, 1 to 2 repetitions, and 120 s being measured once. A resting time of 30 s was given between recording each F-D curve. Each probe-cell was used to test several target-cells. Not more than 40 F-D curves were taken with any given probe-cell. Cells were observed continuously during the SCFS experiment to judge whether they were intact and stably associated with the cantilever/substrate. F-D curves were analysed using IgorPro custom-made routines to extract maximum adhesion force (Figure 2a) and cell deformation during the contact. The F-D curves were pooled and statistically processed as described.

Results

Statistical analysis of SCFS experiments

After determining the maximal adhesion force (Fmax) for each F-D curve, we averaged F*max* over many experimental repetitions to determine the mean adhesion force of a given cell couple and contact time. These Fmax values were pooled to obtain the distribution of maximal adhesion forces for a given experimental condition. The median, mean and standard deviations were extracted with Graph prism (GraphpadSoftware) for unpaired Wilcoxon based Mann-Whitney U-tests for significance with a *p* cut-off value of 0.01. *p* values were calculated using OriginPro (Originlab Inc.) Non-parametric tests were applied to all SCFS data, because we assumed that the data are not normally distributed. Although no systematic history effect was seen on successive F-D curves taken with one cell (data not shown), we could not assume that each curve is strictly independent of each other. Furthermore, cell adhesion forces are likely to be dependent on different properties, e.g. more than one type of adhesion molecule, which does not allow us to use parametric tests. Wilcoxon based Mann-Whitney U-tests are distribution independent and can therefore be applied on composite data sets. Pearson's rank correlation coefficient *R* was computed using Origin Pro.

Protein purification and neurite outgrowth assay

The DNA sequence spanning base pairs 1786-8409 encoding the complete ECD of chicken teneurin-2 was cloned into a modified pSecTag vector (Invitrogen) containing N-terminal His- and FLAG-tags and a 3C protease site and made compatible with the pOPIN suite (Oxford Protein Production Facility; (Berrow et al. 2007) using InFusion cloning (Clontech) and pOPIN compatible primers. HEK293T cells were selected using Zeocin to stably express the ECD. The secreted protein was purified from the conditioned media using Probond resin (Invitrogen) following the manual of the supplier. The purity of our protein preparation is shown in Supplemental Fig. S3A.

Sterile 10 mm cover slips were coated with 30 µg/ml purified protein or 100 µg/ml poly-lysine for at least one hour at room temperature and washed twice with PBS. The cover slips were placed in 4-well tissue culture dishes (Greiner bio-one) and $2x10^4$ cells per dish were seeded in 10% FCS in DMEM. The next day the medium was changed to serum-free DMEM to induce Nb2a differentiation (Evangelopoulos et al. 2005). After 24h cells were fixed with 4% PFA for 7 min at room temperature and stained with rhodamine-coupled

phalloidin (Cytoskeleton Inc.) for 30 min at room temperature and mounted on a glass slide. Pictures were taken on an Olympus IX70 system at 10X magnification and neurite length was manually traced using the NeuronJ (Meijering et al. 2004) plug-in for ImageJ (Abramoff 2004). Only neurites longer than the cell diameter were measured. 100 neurite per condition were measured and statistical significance was tested using the One-Way ANOVA on ranks.

Mixed-culture assay

The mixed-culture assay was performed as described elsewhere in great detail (Biederer and Scheiffele 2007). In short, HEK293-ECR cells expressing either GFP-tagged full-length chicken teneurin-2 (I2E2N2Y2) or GFP- 2E0N0Y0, thus lacking the entire ECD, were co-cultured with dissociated rat hippocampal cells, precultured for 5 days, for additional 2 days followed by fixation and staining for synapsin (Millipore). Five random Z-stacks were acquired per experiment and condition with a Zeiss LSM700 system at 40X magnification and the pictures were deconvolved using the Huygens Remote Manager (Ponti, Schwarb et al. 2007). Co-localization of GFP-tagged teneurin constructs on the HEK293 cells and synapsin of the hippocampal neurites was analyzed using Imaris 7.1.1. (Bitplane). Statistical significance was tested using the One-Way ANOVA on ranks.

Acknowledgments

We thank Hubertus Köhler for his help with the FACS sorting and Steven Bourke and Aaron Ponti for their help in image acquisition and analysis.

Results

References

Abramoff, M. D., Magalhaes, P.J., Ram, S.J. (2004). "**Image Processing with ImageJ**." Biophotonics International 11(7): 36-42.

Bagutti, C., G. Forro, et al. (2003). "**The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription**." Journal of Cell Science 116(Pt 14): 2957-66.

Berrow, N. S., D. Alderton, et al. (2007). "**A versatile ligation-independent cloning method suitable for high-throughput expression screening applications**." Nucleic Acids Research 35(6): e45.

Biederer, T. and P. Scheiffele (2007). "**Mixed-culture assays for analyzing neuronal synapse formation**." Nat. Protocols 2(3): 670-676.

Butt, H. J. and M. Jaschke (1995). "**Calculation of thermal noise in atomic force microscopy**." Nanotechnology 6(1): 1.

Dickson, B. J. (2002). "**Molecular Mechanisms of Axon Guidance**." Science 298(5600): 1959-1964.

Drabikowski, K., A. Trzebiatowska, et al. (2005). "**ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans***." Developmental Biology 282(1): 27-38.

Edwards, T. A., B. D. Wilkinson, et al. (2003). "**Model of the brain tumor-Pumilio translation repressor complex**." Genes Dev 17(20): 2508-2513.

Evangelopoulos, M. E., J. Weis, et al. (2005). "**Signalling pathways leading to neuroblastoma differentiation after serum withdrawal: HDL blocks neuroblastoma differentiation by inhibition of EGFR**." 24(20): 3309-3318.

Feng, K., X.-H. Zhou, et al. (2002). "**All Four Members of the Ten-m/Odz Family of Transmembrane Proteins Form Dimers**." Journal of Biological Chemistry 277(29): 26128-26135.

Fogel, A. I., M. Stagi, et al. (2011). "**Lateral assembly of the immunoglobulin protein SynCAM 1 controls its adhesive function and instructs synapse formation**." EMBO J 30(23): 4728-4738.

Helenius, J., C.-P. Heisenberg, et al. (2008). "**Single-cell force spectroscopy**." Journal of Cell Science 121(11): 1785-1791.

Hong, W., T. J. Mosca, et al. (2012). "**Teneurins instruct synaptic partner matching in an olfactory map**." Nature 484(7393): 201-207.

Janssen, B. J. C., R. A. Robinson, et al. (2010). "**Structural basis of semaphorin-plexin signalling**." Nature 467(7319): 1118-1122.

Kenzelmann, D., R. Chiquet-Ehrismann, et al. (2008). "**Teneurin-1 is expressed in interconnected regions of the developing brain and is processed *in vivo***." BMC Developmental Biology 8: 30.

Kenzelmann, D., R. Chiquet-Ehrismann, et al. (2007). "**Teneurins, a transmembrane protein family involved in cell communication during neuronal development**." Cellular and Molecular Life Sciences 64(12): 1452-1456.

Kurita, S., H. Ogita, et al. (2011). "**Cooperative Role of Nectin-Nectin and Nectin-Afadin Interactions in Formation of Nectin-based Cell-Cell Adhesion**." Journal of Biological Chemistry 286(42): 36297-36303.

Leamey, C. A., S. Merlin, et al. (2007). "**Ten_m3 Regulates Eye-Specific Patterning in the Mammalian Visual Pathway and Is Required for Binocular Vision**." PLoS Biology 5(9): e241.

Liu, H., Z. S. Juo, et al. (2010). "**Structural Basis of Semaphorin-Plexin Recognition and Viral Mimicry from Sema7A and A39R Complexes with PlexinC1**." Cell 142(5): 749-761.

Margadant, C., H. N. Monsuur, et al. (2011). "**Mechanisms of integrin activation and trafficking**." Current Opinion in Cell Biology 23(5): 607-614.

Meijering, E., M. Jacob, et al. (2004). "**Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images**." Cytometry Part A 58A(2): 167-176.

Mörck, C., V. Vivekanand, et al. (2010). "***C. elegans* ten-1 is synthetic lethal with mutations in cytoskeleton regulators, and enhances many axon guidance defective mutants**." BMC Developmental Biology 10(1): 55.

Results

Mosca, T. J., W. Hong, et al. (2012). "**Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice**." Nature 484(7393): 237-241.

Oohashi, T., X.-H. Zhou, et al. (1999). "**Mouse Ten-m/Odz Is a New Family of Dimeric Type II Transmembrane Proteins Expressed in Many Tissues**." Journal of Cell Biology 145(3): 563-577.

Ponti, A., P. Schwarb, et al. (2007). "**Huygens Remote Manager: a Web Interface for High-Volume Batch Deconvolution**." Imaging & Microscopy 9(2): 57-58.

Puech, P.-H., K. Poole, et al. (2006). "**A new technical approach to quantify cell-cell adhesion forces by AFM**." Ultramicroscopy 106(8â€“9): 637-644.

Rubin, B. P., R. P. Tucker, et al. (2002). "**Teneurin 2 is expressed by the neurons of the thalamofugal visual system *in situ* and promotes homophilic cell-cell adhesion *in vitro***." Development 129(20): 4697-705.

Silva, J.-P., V. G. Lelianova, et al. (2011). "**Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities**." Proceedings of the National Academy of Sciences 108(29): 12113-12118.

Tucker, R. P., J. Beckmann, et al. (2011). "**Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry**." Molecular Biology and Evolution.

Tucker, R. P., D. Kenzelmann, et al. (2007). "**Teneurins: Transmembrane proteins with fundamental roles in development**." The International Journal of Biochemistry & Cell Biology 39(2): 292-297.

Young, T. R. and C. A. Leamey (2009). "**Teneurins: Important regulators of neural circuitry**." The International Journal of Biochemistry & Cell Biology 41(5): 990-993.

Zheng, L., Y. Michelson, et al. (2012). "***Drosophila* Ten-m and Filamin Affect Motor Neuron Growth Cone Guidance**." PLoS ONE 6(8): e22956.

Zipursky, S. L. and J. R. Sanes (2010). "**Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly**." Cell 143(3): 343-353.

Results

Figures

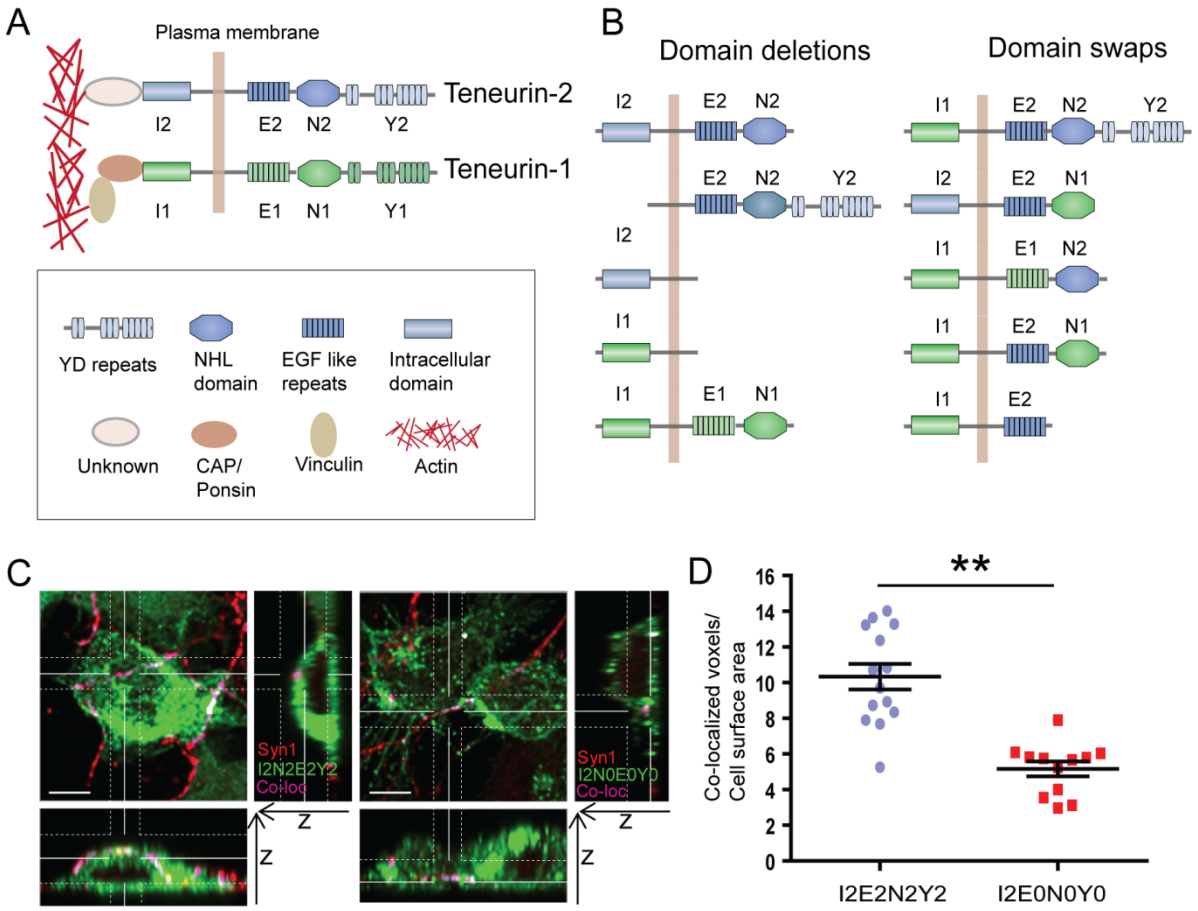


Figure 1: Teneurin structure, constructs used in this study and teneurin expressing cells as a substrate for neurite growth

(A) Domain structure of teneurin-1 and -2 is illustrated. Teneurin-2 domains are in blue and teneurin-1 domains in green. Each domain is denoted with a letter for the domain type followed by the number of the teneurin protein. A possible interaction of the intracellular domains with the actin cytoskeleton is indicated by a link through CAP/ponsin and vinculin or an unknown protein (complex).

(B) Deletion and swap constructs used in this study. The same letter and number code is used as in (A). Missing domains are indicated by their letter code followed by 0.

(C) Co-localization of synapsin with full-length teneurin-2 (I2E2N2Y2) or control cells lacking the ECD (I2E0N0Y0) in a co-culture assay of teneurin expressing HEK293 cells and primary hippocampal neurons. Acquired Z-stacks with slice view in XZ- and YZ-axis are presented. Scale bars, 5 µm.

(D) Quantification of the co-culture assay. The number of co-localized voxels per pictures is normalized for the total cell surface area present in the picture. Black lines represent the median with error bars representing the standard error of the mean (SEM). Statistical significance of differences in the co-localized voxels of three independent experiments is calculated by One-Way ANOVA on ranks (**=p<0.001).

Results

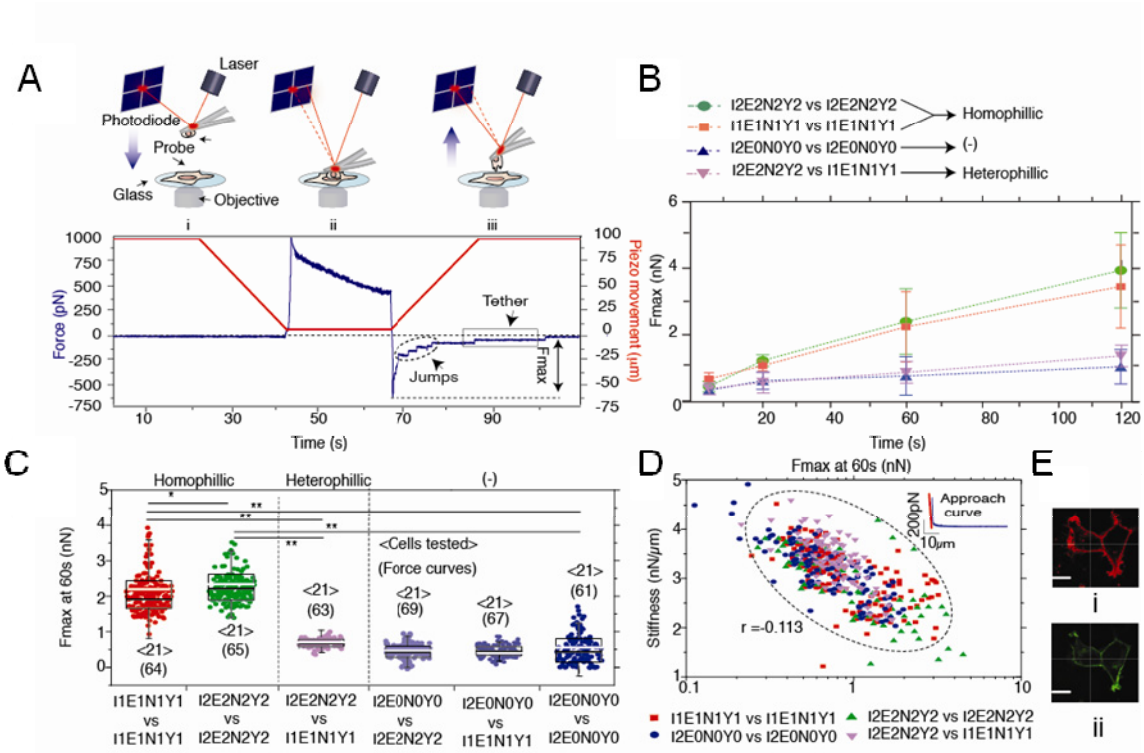


Figure 2: Teneurin expressing cells show contact time dependent adhesion forces upon homophilic interaction measured by SCFS .

(A) Outline of the assay: A single cell immobilized on an AFM cantilever. This 'probe cell' is brought into contact with a second cell adhering to a solid substrate (target cell). After a predefined contact time, the probe cell was retracted at a given speed 5 µm/s and adhesive forces were detected recording the cantilever deflection over the distance travelled by a piezoelectric element moving the cantilever. The resulting force-distance (F-D) curve quantifies the maximum adhesion force (Fmax) between probe and target cell.

(B) Fmax as a function of contact time for homophilic and heterophilic adhesion between three different HEK293 cell lines overexpressing teneurins. Values are presented as mean from 14 to 210 (depending on contact time) F-D curves recorded for each contact time. Error bars are standard errors of the mean (SEM).

(C) Homophilic versus heterophilic adhesion at 60 s contact time. Data is presented as a box-whisker plot with each data point representing one cell-cell adhesion experiment. Median is black and mean is white.

(D) Slope of contact region ('contact stiffness') extracted from the approach F-D curve versus Fmax recorded for each F-D curve. Each cell line is highlighted as separately colored circle, square, upright and inverted triangles. No statistical correlation was seen between the four different cell lines (r = -0.113).

(E) Confocal images of HEK cells expressing teneurins. (i) shows the extracellular domains of teneurin fluorescently stained with an anti-EGF domain antibody. (ii) shows the GFP signal of the N-terminal GFP-tag. Scale bar, 10 µm.

Results

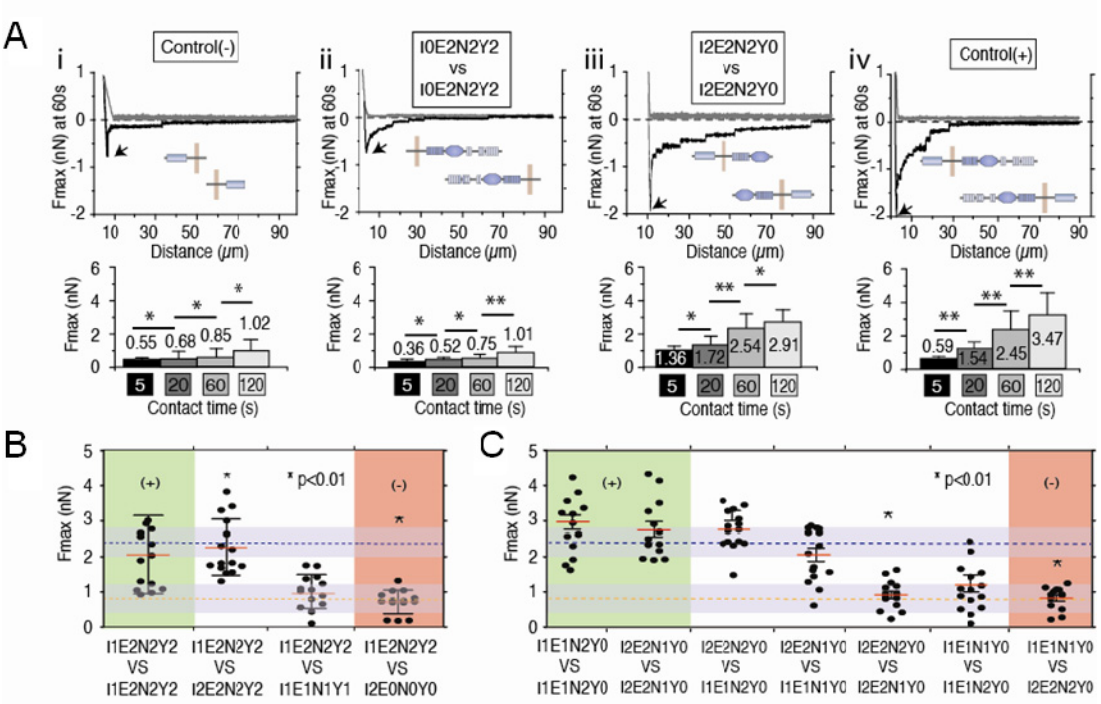


Figure 3: Teneurin ICD and NHL domains, but not the YD repeats, are required for increased adhesion forces measured by SCFS.

(A) Representative F-D curves (60 s contact time) recording the adhesion between HEK cells expressing different teneurin constructs. F-D curves showing the probe cell approaching the target cell are gray. F-D curves recorded when separating both cells (black) are used to estimate the maximum adhesion force (arrow, Fmax). Maximal adhesion forces between probe cell and target cell after contact times of 5 s (black bar), 20 s (dark gray bar), and 60 s (light grey), and 120s (white bar) are presented. 15 to 218 (depending on contact time) F-D curves per condition and contact were acquired. Values are mean ± SEM. Numbers on the bars indicate mean Fmax values. P values are represented as *=p<0.01 and **= p<0.001.

(B, C) Maximal adhesion forces between probe cell and target cell for teneurin constructs in which the ICD, EGF-like repeats and/or NHL domains have been swapped and YD repeats are deleted (see Fig. 1B). Each data point represents Fmax determined for one cell-cell couple where mean adhesion forces are represented as red bars and SEM as black bars. Green and red shaded areas represent positive and negative controls, respectively. The dashed blue line corresponds to mean Fmax of positive controls (homophilic full-length teneurin-1 and -2 adhesion) and the dashed orange line corresponds to mean Fmax of negative controls (heterophilic teneurin-1 and -2 adhesion) (Fig. 3B,C and S3A). Grey shaded area represents SEM of the mean.

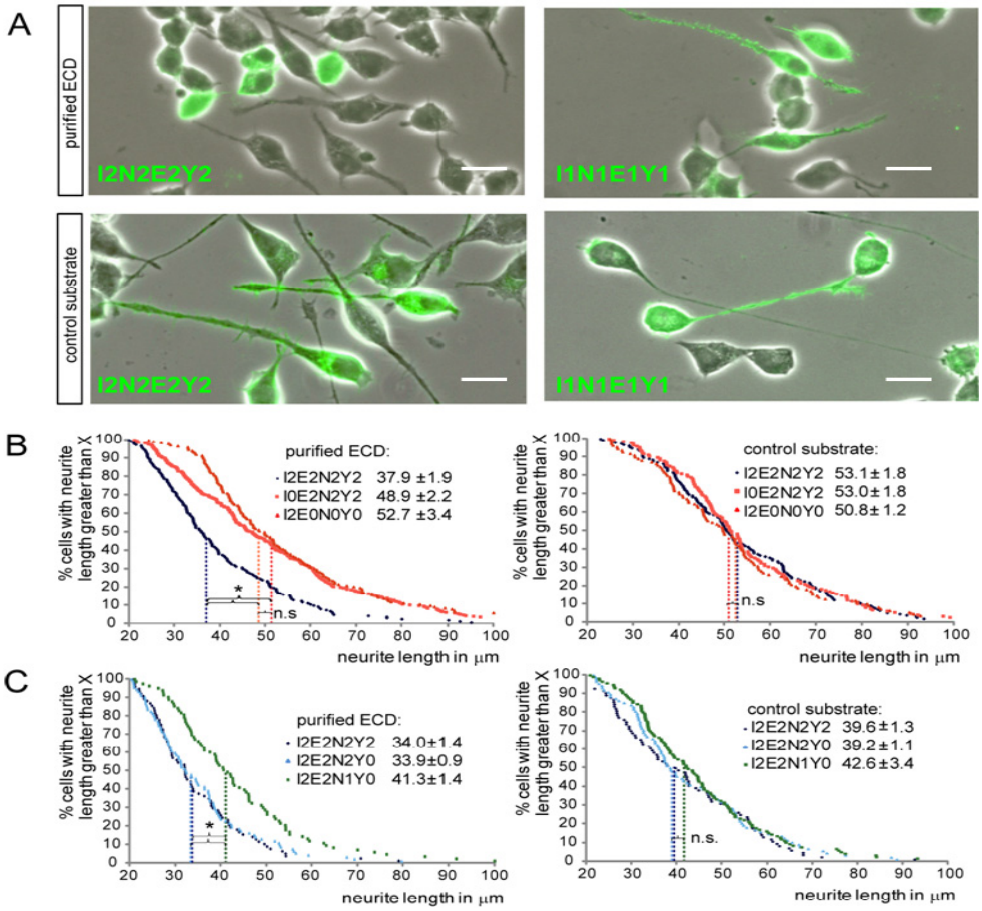


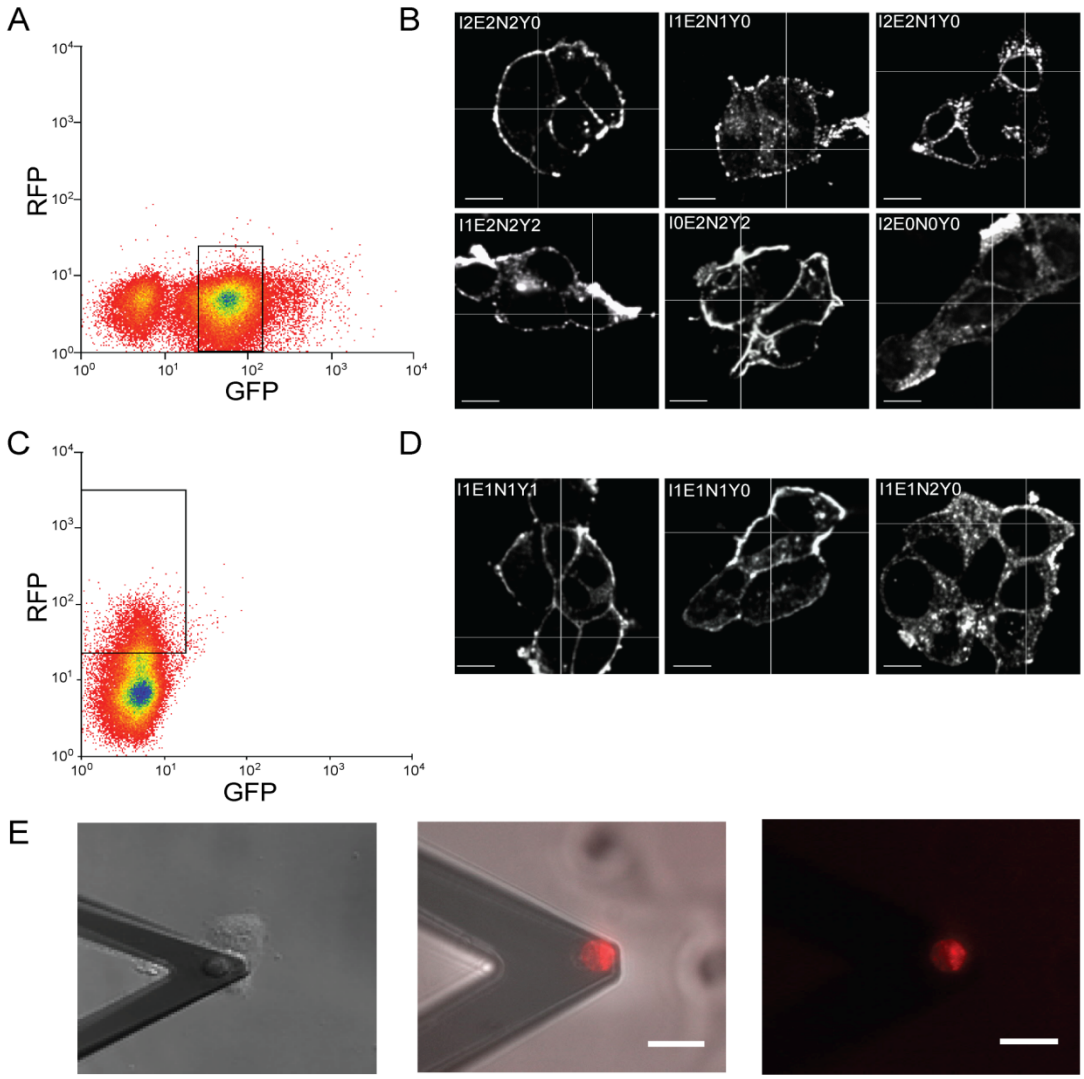
Figure 4: Teneurin homophilic interactions inhibit neurite outgrowth of Nb2a cells mediated by the NHL repeat domain

(A) Nb2a cells transiently transfected with constructs expressing either full-length teneurin-2 (I2E2N2Y2) or teneurin-1 (I1E1N1Y1) were seeded on purified teneurin-2 ECD (top panel) or control substrate (bottom panel). Neurite outgrowth was induced in serum free media. Scale bar, 20 µm.

(B) Quantification results of the neurite outgrowth assay on teneurin-2 ECD (left panel) or poly-lysine (right panel) for Nb2a cells stably expressing full-length teneurin-2 (I2E2N2Y2), a construct with deleted ICD (I0E2N2Y2) or deleted ECD (I2E0N0Y0) are presented. The graph shows the percentage of cells with neurite lengths greater than the length indicated on the X-axis. The horizontal lines represent the median value of three independent experiments and the median values and SEM for each construct are indicated next to the legend. Significance of differences in median values is calculated by One-Way ANOVA on ranks (*=p<0.01).

(D) Quantification results of the neurite outgrowth assay on teneurin-2 ECD (left panel) or poly-lysine (right panel) of Nb2a cells stably expressing full-length teneurin-2 (I2N2E2Y2), a construct with deleted YD repeats (I2E2N2Y0) or with the NHL domain swapped (I2E2N1Y0) are presented. The graph shows the percentage of cells with neurite lengths greater than the length indicated on the X-axis. The horizontal lines represent the median value of three independent experiments and the median values and SEM for each construct are indicated next to the legend. Significance of differences in median values is calculated by One-Way ANOVA on ranks (*=p<0.01).

Results



Supplementary Figure S1: Expression analysis with FACS, teneurin surface staining and cell attached to the cantilever for SCFS

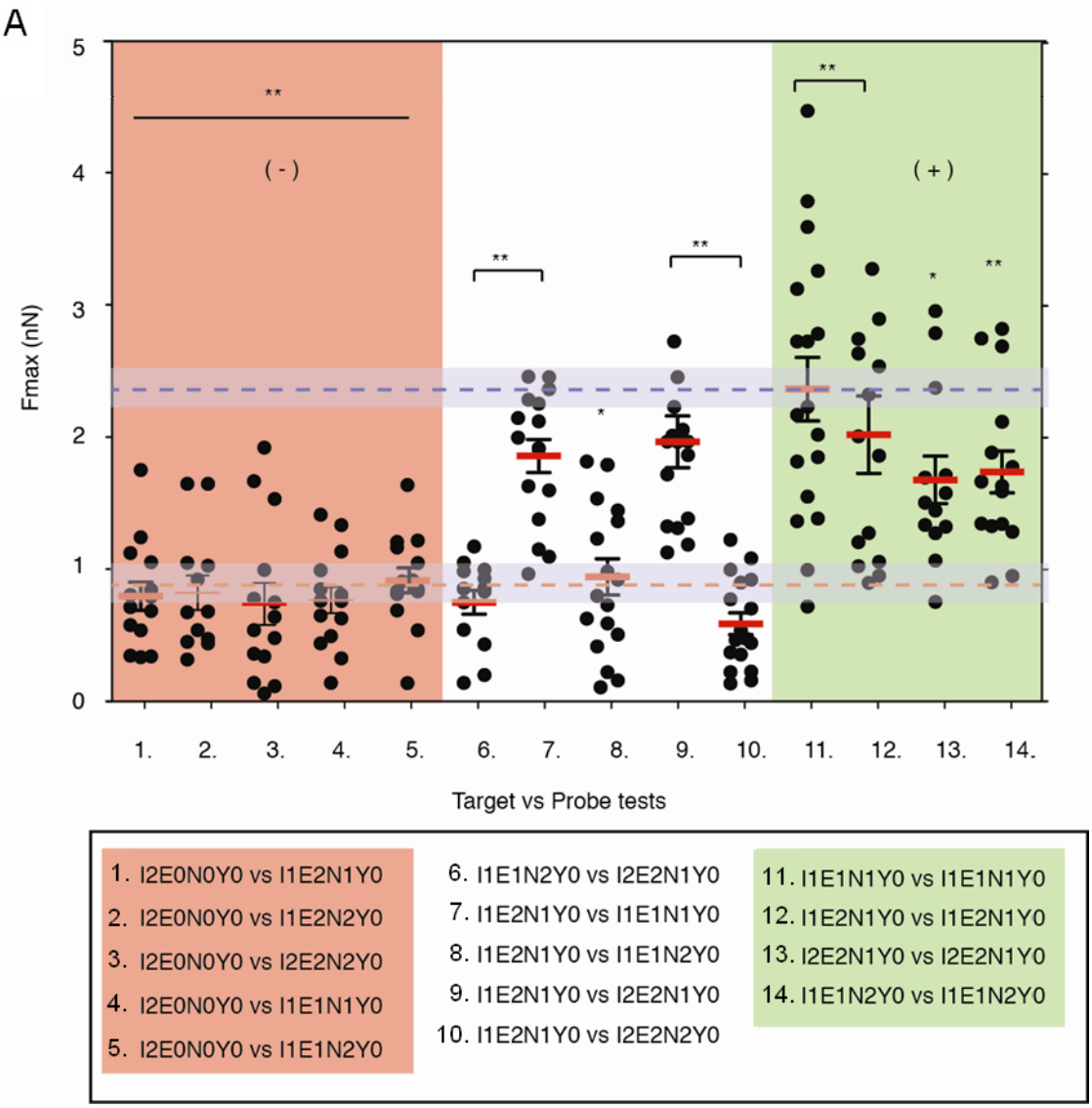
(A) Representative FACS profile for HEK293 cells expressing full-length GFP-tagged teneurin-2. The box represents the gate for sorting of all GFP-expressing cell lines.

(B) Surface stainings with an anti-EGF teneurin-2 antibody of non-permeabilized HEK293 cells expressing the constructs indicated and sorted with the gating presented in (A). Scale bar, 10 µm.

(C) Representative FACS profile for HEK293 cells expressing full-length RFP-tagged teneurin-1. The box represents the gate for sorting of all RFP-expressing cell lines.

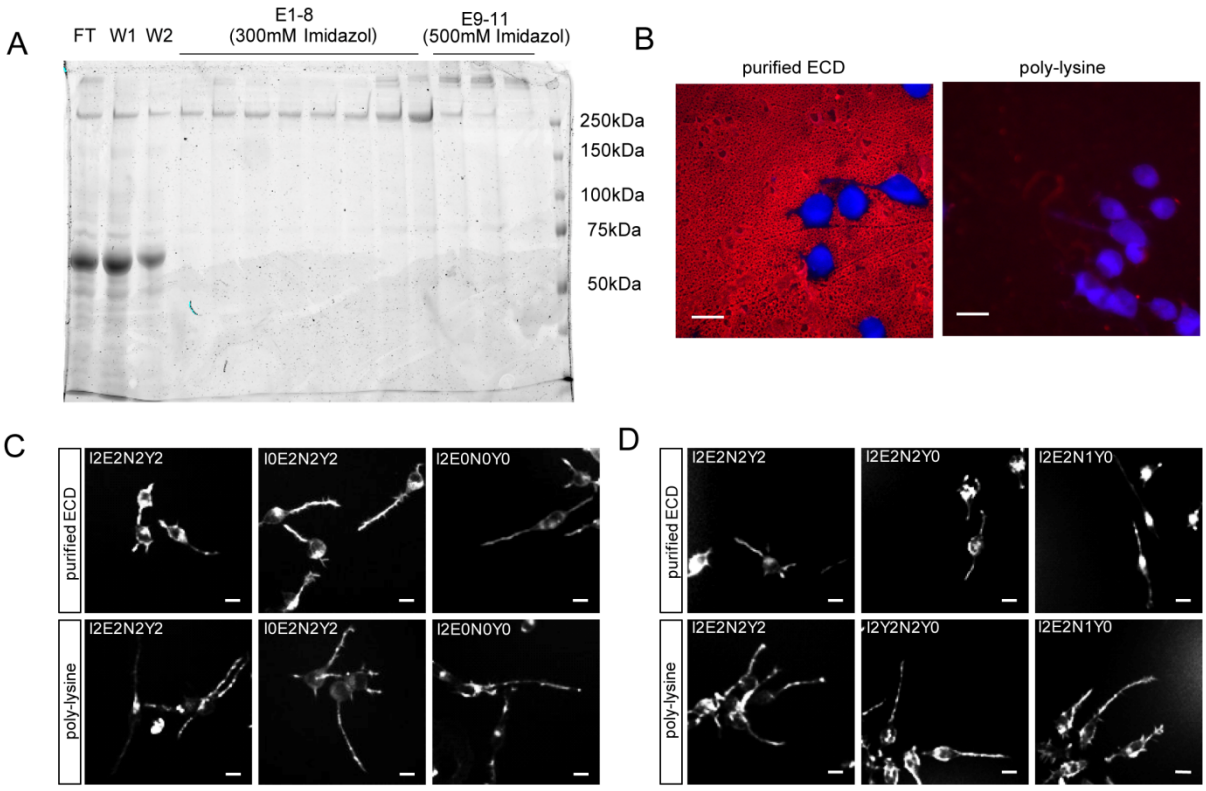
(D) Surface stainings with an anti-EGF teneurin-1 antibody of non-permeabilized HEK293 cells expressing the constructs indicated and sorted with the gating presented in (C). Scale bar, 10 µm.

(E) Representative image of the loading of the cantilever with a probe cell for SCFS. The cantilever is brought on top of a cell and picked up and cell fluorescence is checked. Scale bar, 20 µm.



Supplementary Figure S2: Contributions of various teneurin constructs on cell adhesion.

(A) Presented are the maximal adhesion force measurements between HEK293 cells expressing teneurin swap and deletion constructs. These data complement the data shown in Figure 3C. The red and green shaded areas show constructs that are considered as negative and positive controls respectively. The blue dashed line indicates mean Fmax of homophilic adhesion between teneurins (positive control) and the orange dashed line indicates mean Fmax of heterophilic adhesion between teneurins (negative control) (Fig. 2C) with the gray shaded area representing the SEM. Significance was calculated by measuring the significance of each data set in comparison to the blue dashed line. **=p<0.001; *=p<0.01



Supplementary Figure S3: Purity of protein purification, coating efficiency of cover slips and representative images of the neurite outgrowth assay.

(A) A coomassie stained protein gel of the protein purification is shown. Elution fractions (E) 1-8 show the purified protein migrating at the size of monomeric protein of about 250kDa of size. These fractions were pooled, confirmed for teneurin protein identity and purity in mass spectrometry (data not shown) and used for coating of cover slips. (FT) flow through; (W) wash fraction; (E) elution fraction. E9-11 contain non-reduced dimers of the protein.

(B) Glass cover slips were coated with purified protein of (A) and Nb2a cells were seeded on top of them. Coating efficiency was visualized by staining with an anti-EGF of teneurin-2 antibody and cell nuclei are DAPI stained. Scale bars, 10 µm.

(C, D) Representative pictures of the neurite outgrowth assays on glass cover slips coated with purified ECD as shown in (B) or on poly-lysine as a control substrate (presented in Figure 4B,C). Scale bar, 10 µm.

5. Conclusions

The results obtained in the different studies presented above are discussed in detail in each of the publications. Therefore, I will only add a small section to describe how my findings are related to each other and how they help to improve our understanding of teneurin functions. The study presented in chapter 4.1. "Phylogenetic Analysis of the Teneurins: Conserved Features and Premetazoan Ancestry" describes teneurins as a phylogenetically old and conserved protein family. We found a teneurin-like gene in the choanoflagellate *Monosiga breviocollis*, an unicellular eukaryote considered to be the closest relative of the metazoan.[100, 101] In the study "Teneurin homophilic interaction is mediated by their propeller domain and inhibits neurite outgrowth" as presented in chapter 4.3., we show that teneurins homophilically interact to create strong adhesion forces between cells. This function may indicate that the presence of a teneurin-like gene in a non-metazoan organism is involved in facilitating metazoan multicellularity, similar to the proposed role of cadherins in this process.[102] Furthermore, the domain structures between species and paralogs were analyzed in more detail in the first study. Great emphasis was placed on the ICD, since it is believed to play an important role in mediating teneurin functions. Previously, it was shown that the teneurin-1 ICD interacts with the adaptor protein CAP/ponsin, representing a possible link to the cytoskeleton. Even though the known SH3 binding motif only differs in one amino acid in teneurin-2, it is not able to bind CAP/ponsin.[80] The results obtained in the third study presented, clearly show the importance of the teneurin ICD in mediating adhesion forces between two cells. Thus, a connection between the ICD and the cytoskeleton seems to be crucial. We could show that this feature is conserved between the ICDs of teneurin-1 and -2, since ICD swap constructs were able to mediate the same adhesion forces as the original constructs. As teneurin-2 is not able to bind CAP/ponsin, there has to be another conserved link between teneurin ICDs and the cytoskeleton. CAP/ponsin is part of the vinexin family of adaptor proteins. All three family members have been shown to perform a function in cell adhesion.[103] It would be interesting to analyze if different teneurins bind to different members of the vinexin protein family. In addition, a screen for further interaction partners for all teneurin ICDs would help to better understand how the ICD is able to mediate teneurin functions.

Axon guidance molecules as well as target finding molecules usually have common characteristics. They promote adhesion and neurite outgrowth and are expressed in distinct

patterns. Extensive studies of teneurin expression in different species revealed distinct, non-overlapping expression patterns. These patterns are usually regulated by a network of transcription factors. In the study “Human Teneurin-1 is a direct target of the homeobox transcription factor EMX2 at a novel alternate promoter” as presented in section 4.2., we show that the expression of teneurin-1 is directly regulated by Emx2, a prominent member of this transcription factor network. Emx2 directly binds to a single homeobox binding site in the novel alternate promoter in both, *in vitro* and *in vivo*. This regulation might contribute to the patterned teneurin-1 expression in the developing brain. In addition, as shown in chapter 4.3., teneurins mediate homophilic adhesion forces, but not heterophilic ones. The domain responsible for this discrimination is the propeller-like NHL domain. Furthermore, an overexpression of teneurin-2 leads to neurite outgrowth,[81] which is slowed down when teneurins interact homophilically.

In conclusion, the results presented here, together with increasing evidence in literature, clearly indicate that teneurins are axon guidance and target finding molecules. This function is phylogenetically conserved. Finally, the ICD and the NHL propeller domain were identified as the important domains in mediating this function. This provides the basis to further analyze these domains. One major question to be answered in future studies is which proteins are involved in linking the ICD to the cytoskeleton. Furthermore, it will be interesting to examine how the NHL domains of teneurin-1 and -2 can discriminate between homophilic and heterophilic binding, despite their high conservation. Solving the crystal structure of this domain to get information about which residues are crucial in discriminating homophilic interactions would be a good approach to understand the function of the NHL domains in more detail.

6. References

1. Lodish H., et al., **Molecular Cell Biology, 5th Edition**. 2003: W.H. Freeman and Company, New York, NY, USA.
2. Vasioukhin, V. and E. Fuchs, **Actin dynamics and cell-cell adhesion in epithelia**. Current Opinion in Cell Biology, 2001. 13(1): p. 76-84.
3. Redies, C., F. Neudert, and J. Lin, **Cadherins in Cerebellar Development: Translation of Embryonic Patterning into Mature Functional Compartmentalization**. The Cerebellum, 2011. 10(3): p. 393-408.
4. Hynes, R.O. and A.D. Lander, **Contact and adhesive specificities in the associations, migrations, and targeting of cells and axons**. Cell, 1992. 68(2): p. 303-322.
5. Hansen, S., V. Berezin, and E. Bock, **Signaling mechanisms of neurite outgrowth induced by the cell adhesion molecules NCAM and N-Cadherin**. Cellular and Molecular Life Sciences, 2008. 65(23): p. 3809-3821.
6. Walmod, P.S., et al., **Zippers Make Signals: NCAM-mediated Molecular Interactions and Signal Transduction**. Neurochemical Research, 2004. 29(11): p. 2015-2035.
7. Margadant, C., et al., **Mechanisms of integrin activation and trafficking**. Current Opinion in Cell Biology, 2011. 23(5): p. 607-614.
8. Stiles, J. and T. Jernigan, **The Basics of Brain Development**. Neuropsychology Review, 2010. 20(4): p. 327-348.
9. Stiles, J., **The fundamentals of brain development. Integrating nature and nurture**. 2008: Harvard University Press, Cambridge, MA, USA.
10. Sanes, D., T. Reh, and W. Harris, **Development of the nervous system. 2nd ed**. 2000, San Diego: Academic Press.
11. Sur, M. and J.L.R. Rubenstein, **Patterning and Plasticity of the Cerebral Cortex**. Science, 2005. 310(5749): p. 805-810.
12. O'Leary, D.D.M., S.-J. Chou, and S. Sahara, **Area Patterning of the Mammalian Cortex**. Neuron, 2007. 56(2): p. 252-269.
13. Sansom, S.N. and F.J. Livesey, **Gradients in the brain: the control of the development of form and function in the cerebral cortex**. Cold Spring Harb Perspect Biol, 2009. 1(2): p. a002519.
14. Lumsden, A. and R. Krumlauf, **Patterning the Vertebrate Neuraxis**. Science, 1996. 274(5290): p. 1109-1115.
15. Keynes, R. and R. Krumlauf, **Hox Genes and Regionalization of the Nervous System**. Annual Review of Neuroscience, 1994. 17(1): p. 109-132.
16. Crossley, P.H., et al., **Coordinate expression of Fgf8, Otx2, Bmp4, and Shh in the rostral prosencephalon during development of the telencephalic and optic vesicles**. Neuroscience, 2001. 108(2): p. 183-206.
17. Furuta, Y., D.W. Piston, and B.L. Hogan, **Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development**. Development, 1997. 124(11): p. 2203-2212.
18. Crossley, P.H. and G.R. Martin, **The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo**. Development, 1995. 121(2): p. 439-451.
19. Ohkubo, Y., C. Chiang, and J.L.R. Rubenstein, **Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles**. Neuroscience, 2002. 111(1): p. 1-17.

Appendix

20. Rubenstein, J.L.R., et al., **Regionalization of the prosencephalic neural plate**. Annual Review of Neuroscience, 1998. 21(1): p. 445-477.
21. Bishop, K.M., G. Goudreau, and D.D.M. O'Leary, **Regulation of Area Identity in the Mammalian Neocortex by Emx2 and Pax6**. Science, 2000. 288(5464): p. 344-349.
22. Osumi, N., **The Role of Pax6 in Brain Patterning**. The Tohoku Journal of Experimental Medicine, 2001. 193(3): p. 163-174.
23. Bishop, K.M., J.L.R. Rubenstein, and D.D.M. O'Leary, **Distinct Actions of Emx1, Emx2, and Pax6 in Regulating the Specification of Areas in the Developing Neocortex**. The Journal of Neuroscience, 2002. 22(17): p. 7627-7638.
24. Hamasaki, T., et al., **EMX2 Regulates Sizes and Positioning of the Primary Sensory and Motor Areas in Neocortex by Direct Specification of Cortical Progenitors**. Neuron, 2004. 43(3): p. 359-372.
25. Tessier-Lavigne, M. and C.S. Goodman, **The Molecular Biology of Axon Guidance**. Science, 1996. 274(5290): p. 1123-1133.
26. Dickson, B.J., **Molecular Mechanisms of Axon Guidance**. Science, 2002. 298(5600): p. 1959-1964.
27. Kidd, T., K.S. Bland, and C.S. Goodman, **Slit Is the Midline Repellent for the Robo Receptor in Drosophila**. Cell, 1999. 96(6): p. 785-794.
28. Battye, R., A. Stevens, and J.R. Jacobs, **Axon repulsion from the midline of the Drosophila CNS requires slit function**. Development, 1999. 126(11): p. 2475-2481.
29. Plump, A.S., et al., **Slit1 and Slit2 Cooperate to Prevent Premature Midline Crossing of Retinal Axons in the Mouse Visual System**. Neuron, 2002. 33(2): p. 219-232.
30. Culotti, J.G. and D.C. Merz, **DCC and netrins**. Current Opinion in Cell Biology, 1998. 10(5): p. 609-613.
31. Yee, K.T., et al., **Extension of Long Leading Processes and Neuronal Migration in the Mammalian Brain Directed by the Chemoattractant Netrin-1**. Neuron, 1999. 24(3): p. 607-622.
32. Sun, K.L.W., J.P. Correia, and T.E. Kennedy, **Netrins: versatile extracellular cues with diverse functions**. Development, 2011. 138(11): p. 2153-2169.
33. Yazdani, U. and J. Terman, **The semaphorins**. Genome Biology, 2006. 7(3): p. 211.
34. Raper, J.A., **Semaphorins and their receptors in vertebrates and invertebrates**. Current Opinion in Neurobiology, 2000. 10(1): p. 88-94.
35. Wilkinson, D.G., **Multiple roles of eph receptors and ephrins in neural development**. Nat Rev Neurosci, 2001. 2(3): p. 155-164.
36. Frisén, J., et al., **Ephrin-A5 (AL-1/RAGS) Is Essential for Proper Retinal Axon Guidance and Topographic Mapping in the Mammalian Visual System**. Neuron, 1998. 20(2): p. 235-243.
37. Ditlevsen, D.K., K. Kolkova, and V. Berezin, **Signaling Pathways Involved in NCAM-Induced Neurite Outgrowth: Structure and Function of the Neural Cell Adhesion Molecule NCAM**. Advances in Experimental Medicine and Biology, 2010. 663: p. 151-168.
38. Gundersen, R.W., **Response of sensory neurites and growth cones to patterned substrata of laminin and fibronectin in vitro**. Developmental Biology, 1987. 121(2): p. 423-431.
39. Faissner, A., **The tenascin gene family in axon growth and guidance**. Cell and Tissue Research, 1997. 290(2): p. 331-341.
40. Waites, C.L., A.M. Craig, and C.C. Garner, **Mechanisms of vertebrate synaptogenesis**. Annual Review of Neuroscience, 2005. 28(1): p. 251-274.

Appendix

41. Scheiffele, P., **Cell-Cell signaling during synapse formation in the CNS**. Annual Review of Neuroscience, 2003. 26(1): p. 485-508.
42. Shapiro, L. and D.R. Colman, **The Diversity of Cadherins and Implications for a Synaptic Adhesive Code in the CNS**. Neuron, 1999. 23(3): p. 427-430.
43. Inoue, A. and J.R. Sanes, **Lamina-Specific Connectivity in the Brain: Regulation by N-Cadherin, Neurotrophins, and Glycoconjugates**. Science, 1997. 276(5317): p. 1428-1431.
44. Lee, C.-H., et al., **N-Cadherin Regulates Target Specificity in the Drosophila Visual System**. Neuron, 2001. 30(2): p. 437-450.
45. Lee, R.C., et al., **The protocadherin Flamingo is required for axon target selection in the Drosophila visual system**. Nat Neurosci, 2003. 6(6): p. 557-563.
46. O`Brien, R.J., et al., **Synaptic Clustering of AMPA Receptors by the Extracellular Immediate-Early Gene Product Narp**. Neuron, 1999. 23(2): p. 309-323.
47. Dalva, M.B., et al., **EphB Receptors Interact with NMDA Receptors and Regulate Excitatory Synapse Formation**. Cell, 2000. 103(6): p. 945-956.
48. Biederer, T., et al., **SynCAM, a Synaptic Adhesion Molecule That Drives Synapse Assembly**. Science, 2002. 297(5586): p. 1525-1531.
49. Akins, M.R. and T. Biederer, **Cell-cell interactions in synaptogenesis**. Current Opinion in Neurobiology, 2006. 16(1): p. 83-89.
50. Minet, A.D. and R. Chiquet-Ehrismann, **Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli***. Gene, 2000. 257(1): p. 87-97.
51. Baumgartner, S. and R. Chiquet-Ehrismann, **Tena, a *Drosophila* gene related to tenascin, shows selective transcript localization.** Mechanisms of Development, 1993(40): p. 165-176.
52. Baumgartner, S., et al., **Ten(m), a *Drosophila* gene related to Tenascin, is a new pair-rule gene**. EMBO Journal, 1994. 13(16): p. 3728-3740.
53. Levine, A., et al., ***odd Oz*: A novel *Drosophila* pair rule gene**. Cell, 1994. 77(4): p. 587-598.
54. Zheng, L., et al., ***Drosophila* Ten-m and Filamin Affect Motor Neuron Growth Cone Guidance**. PLoS ONE, 2012. 6(8): p. e22956.
55. Drabikowski, K., A. Trzebiatowska, and R. Chiquet-Ehrismann, **ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans***. Developmental Biology, 2005. 282(1): p. 27-38.
56. Minet, A.D., et al., **Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain**. Journal of Cell Science, 1999. 112 (Pt 12): p. 2019-32.
57. Mörck, C., et al., ***C. elegans* ten-1 is synthetic lethal with mutations in cytoskeleton regulators, and enhances many axon guidance defective mutants**. BMC Developmental Biology, 2010. 10(1): p. 55.
58. Levine, A., C. Weiss, and R. Wides, **Expression of the pair-rule gene odd Oz (odz) in imaginal tissues**. Developmental Dynamics, 1997. 209(1): p. 1-14.
59. Fascetti, N. and S. Baumgartner, **Expression of Drosophila Ten-a, a dimeric receptor during embryonic development**. Mechanisms of Development, 2002. 114(1â€“2): p. 197-200.

60. Mieda, M., et al., **Compartmentalized expression of zebrafish ten-m3 and ten-m4, homologues of the Drosophila tenm/odd Oz gene, in the central nervous system**. Mechanisms of Development, 1999. 87(1-2): p. 223-227.
61. Rubin, B.P., et al., **Teneurins: A Novel Family of Neuronal Cell Surface Proteins in Vertebrates, Homologous to the *Drosophila* Pair-Rule Gene Product Ten-m**. Developmental Biology, 1999. 216(1): p. 195-209.
62. Kenzelmann, D., et al., **Teneurin-1 is expressed in interconnected regions of the developing brain and is processed *in vivo***. BMC Developmental Biology, 2008. 8: p. 30.
63. Kenzelmann Broz, D., Tucker R.P., Leachman N.T. and Chiquet-Ehrismann R., **The expression of teneurin-4 in the avian embryo: potential roles in patterning of the limb and nervous system**. Int. J. Dev. Biol., 2010. 54: p. 1509-1516.
64. Tucker, R.P., et al., **Teneurin-2 is expressed in tissues that regulate limb and somite pattern formation and is induced *in vitro* and *in situ* by FGF8**. Developmental Dynamics, 2001. 220(1): p. 27-39.
65. Tucker, R.P., et al., **The expression of teneurin-4 in the avian embryo**. Mechanisms of Development, 2000. 98(1-2): p. 187-91.
66. Otaki, J.M. and S. Firestein, **Segregated expression of neurestin in the developing olfactory bulb**. Neuroreport, 1999. 10(12): p. 2677-2680.
67. Oohashi, T., et al., **Mouse Ten-m/Odz Is a New Family of Dimeric Type II Transmembrane Proteins Expressed in Many Tissues**. Journal of Cell Biology, 1999. 145(3): p. 563-577.
68. Ben-Zur, T., et al., **The Mammalian Odz Gene Family: Homologs of a *Drosophila* Pair-Rule Gene with Expression Implying Distinct yet Overlapping Developmental Roles**. Developmental Biology, 2000. 217(1): p. 107-120.
69. Li, H., K.M. Bishop, and D.D. O'Leary, **Potential target genes of EMX2 include Odz/Ten-M and other gene families with implications for cortical patterning**. Molecular and Cellular Neuroscience, 2006. 33(2): p. 136-49.
70. Zhou, X.-H., et al., **The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain**. Gene Expression Patterns, 2003. 3(4): p. 397-405.
71. Ben-Zur, T. and R. Wides, **Mapping Homologs of *Drosophila* odd Oz(odz):Doc4/Odz4 to Mouse Chromosome 7, Odz1to Mouse Chromosome 11 and ODZ3 to Human Chromosome Xq25**. Genomics, 1999. 58(1): p. 102-103.
72. Lossie, A.C., et al., **Mutation of l7Rn3 Shows That Odz4 Is Required for Mouse Gastrulation**. Genetics, 2005. 169(1): p. 285-299.
73. Wang, X.-Z., et al., **Identification of novel stress-induced genes downstream of chop**. EMBO Journal, 1998. 17(13): p. 3619-3630.
74. Tucker, R.P., et al., **Teneurins: Transmembrane proteins with fundamental roles in development**. The International Journal of Biochemistry & Cell Biology, 2007. 39(2): p. 292-297.
75. Tucker, R.P. and R. Chiquet-Ehrismann, **Teneurins: A conserved family of transmembrane proteins involved in intercellular signaling during development**. Developmental Biology, 2006. 290(2): p. 237-245.
76. Feng, K., et al., **All Four Members of the Ten-m/Odz Family of Transmembrane Proteins Form Dimers**. Journal of Biological Chemistry, 2002. 277(29): p. 26128-26135.

Appendix

77. Edwards, T.A., et al., **Model of the brain tumor-Pumilio translation repressor complex**. Genes Dev, 2003. 17(20): p. 2508-2513.
78. Lovejoy, D.A., S. Rotzinger, and D. Barsyte-Lovejoy, **Evolution of Complementary Peptide Systems: teneurin C-terminal-associated peptides and corticotropin-releasing factor superfamilies**. Annals of the New York Academy of Sciences, 2009. 1163(Trends in Comparative Endocrinology and Neurobiology): p. 215-220.
79. Trzebiatowska, A., et al., ***Caenorhabditis elegans*** **teneurin, ten-1, is required for gonadal and pharyngeal basement membrane integrity and acts redundantly with integrin ina-1 and dystroglycan dgn-1**. Molecular Biology of the Cell, 2008. 19(9): p. 3898-908.
80. Nunes, S.M., et al., **The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix**. Experimental Cell Research, 2005. 305(1): p. 122-32.
81. Rubin, B.P., et al., **Teneurin 2 is expressed by the neurons of the thalamofugal visual system *in situ* and promotes homophilic cell-cell adhesion *in vitro***. Development, 2002. 129(20): p. 4697-705.
82. Bagutti, C., et al., **The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription**. Journal of Cell Science, 2003. 116(Pt 14): p. 2957-66.
83. Brown, M.S., et al., **Regulated Intramembrane Proteolysis: A Control Mechanism Conserved from Bacteria to Humans**. Cell, 2000. 100(4): p. 391-398.
84. Kenzelmann, D., R. Chiquet-Ehrismann, and R. Tucker, **Teneurins, a transmembrane protein family involved in cell communication during neuronal development**. Cellular and Molecular Life Sciences, 2007. 64(12): p. 1452-1456.
85. Wade, P.A., **Methyl CpG-binding proteins and transcriptional repression***. BioEssays, 2001. 23(12): p. 1131-1137.
86. Topf, U. and R. Chiquet-Ehrismann, **Genetic interaction between *Caenorhabditis elegans* teneurin ten-1 and prolyl 4-hydroxylase phy-1 and their function in collagen IV-mediated basement membrane integrity during late elongation of the embryo**. Molecular Biology of the Cell, 2011. 22(18): p. 3331-3343.
87. Silva, J.-P., et al., **Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities**. Proceedings of the National Academy of Sciences, 2011. 108(29): p. 12113-12118.
88. Kinel-Tahan, Y., et al., ***Drosophila*** **odz gene is required for multiple cell types in the compound retina**. Developmental Dynamics, 2007. 236(9): p. 2541-2554.
89. Leamey, C.A., et al., **Differential Gene Expression between Sensory Neocortical Areas: Potential Roles for Ten_m3 and Bcl6 in Patterning Visual and Somatosensory Pathways**. Cerebral Cortex, 2008. 18(1): p. 53-66.
90. Leamey, C.A., et al., **Ten_m3 Regulates Eye-Specific Patterning in the Mammalian Visual Pathway and Is Required for Binocular Vision**. PLoS Biology, 2007. 5(9): p. e241.
91. Luckasson, R. and A. Reeve, **Naming, Defining, and Classifying in Mental Retardation**. Mental Retardation, 2001. 39(1): p. 47-52.
92. Plenge, R.M., et al., **Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders**. The American Journal of Human Genetics, 2002. 71(1): p. 168-173.
93. Gécz, J., C. Shoubridge, and M. Corbett, **The genetic landscape of intellectual disability arising from chromosome X**. Trends in Genetics, 2009. 25(7): p. 308-316.

Appendix

94. Chiurazzi, P., et al., **XLMR genes: update 2007**. Eur J Hum Genet, 2008. 16(4): p. 422-434.
95. Chelly, J., et al., **Genetics and pathophysiology of mental retardation**. Eur J Hum Genet, 2006. 14(6): p. 701-713.
96. Raymond, F.L., et al., **Mutations in ZDHHC9, Which Encodes a Palmitoyltransferase of NRAS and HRAS, Cause X-Linked Mental Retardation Associated with a Marfanoid Habitus**. The American Journal of Human Genetics, 2007. 80(5): p. 982-987.
97. Chiyonobu, T., et al., **Partial tandem duplication of GRIA3 in a male with mental retardation**. American Journal of Medical Genetics Part A, 2007. 143A(13): p. 1448-1455.
98. Tarpey, P.S., et al., **Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation**. Nat Genet, 2007. 39(9): p. 1127-1133.
99. Li, H., et al., **Mutation analysis of methyl-CpG binding protein family genes in autistic patients**. Brain and Development, 2005. 27(5): p. 321-325.
100. Lang, B.F., et al., **The Closest Unicellular Relatives of Animals**. Current Biology, 2002. 12(20): p. 1773-1778.
101. Ruiz-Trillo, I.a., et al., **A Phylogenomic Investigation into the Origin of Metazoa**. Molecular Biology and Evolution, 2008. 25(4): p. 664-672.
102. Abedin, M. and N. King, **The Premetazoan Ancestry of Cadherins**. Science, 2008. 319(5865): p. 946-948.
103. Kioka, N., K. Ueda, and T. Amachi, **Vinexin, CAP/ponsin, ArgBP2: a Novel Adaptor Protein Family Regulating Cytoskeletal Organization and Signal Transduction**. Cell Structure and Function, 2002. 27(1): p. 1-7.

7. Appendix

7.1. List of Figures and Tables

7.2. Abbreviations

AFM	atomic force microscopy
BMP	bone morphogenic protein
CAM	cell adhesion molecule
CNS	central nervous system
CRF	corticotrophin releasing factor
DCC	deleted in colorectal carcinoma
ECD	extracellular domain
ECM	extracellular matrix
EGF-like	epidermal growth factor-like
Eph	ephrin
Fgf	fibroblast growth factor
ICD	intracellular domain
Ig	immunoglobular
MR	mental retardation
RIP	regulated intramembrane proteolysis
SCFS	single cell force spectroscopy
Shh	sonic hedgehog
ten-a	tenascin-like accessory
ten-m	tenascin-like major
TCAP	teneurin C-terminal associated peptide
XLMR	X-linked mental retardation

7.3. Acknowledgements

First of all, I want to thank Dr. Ruth Chiquet-Ehrismann to give me the opportunity to do my PhD work in her lab and the FMI. I am very grateful to her for her guidance, her always open door and discussions over my project and for giving me the chance to perform my work independently and bringing in my own ideas.

I would like to thank my thesis committee members Dr. Jan Pielage for his scientific input in the project and Richard P. Tucker, PhD for many discussions and sharing his great knowledge of teneurins and for revising the manuscripts of my publications.

Furthermore, I like to thank Jacqueline Ferralli for her never ending support with all kinds of experiments and for the chats about all aspects of laboratory- and institute-life and of course sharing her strong opinion about all sport events.

I would like to thank all present and past members of the Chiquet lab for the really nice working athmosphere and the daily lunch breaks (some more funny than others). I especially want to thank Matthias Scharenberg, Jonas Schöler and Claudius Kratochwil for the daily coffee breaks (some more on time than others) and for the extensive discussions about science, sports and everything else and for being my friends. And I want to thank Enrico Martina, my fellow senior PhD student for many valuable inputs and always sharing his funny stories (again some more funny than others).

Furthermore, I want to thank my and Susis whole family for their ongoing support, but especially my parents Christhild and Hans-Jürgen Beckmann for always being there for me, supporting me in my decisions, even though it means to move quite a distant away from them and for contributing a lot for me to become the person I am now.

I like to thank all my friends -near and far- for doing all those things friends are for.

Finally, I am most grateful to Susi Finger for being there for me in all aspects of life and the countless things she has done and still does for me.